

Scotland's Rural College

Analysis of genome-wide DNA arrays reveals the genomic population structure and diversity in autochthonous Greek goat breeds

Michailidou, S.; Tsangaris, G. Th; Tzora, A.; Skoufos, I.; Banos, G.; Argiriou, A.; Arsenos, G.

Published in:
PLOS ONE

DOI:
[10.1371/journal.pone.0226179](https://doi.org/10.1371/journal.pone.0226179)

First published: 12/12/2019

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (APA):

Michailidou, S., Tsangaris, G. T., Tzora, A., Skoufos, I., Banos, G., Argiriou, A., & Arsenos, G. (2019). Analysis of genome-wide DNA arrays reveals the genomic population structure and diversity in autochthonous Greek goat breeds. *PLoS ONE*, 14(12), [e0226179]. <https://doi.org/10.1371/journal.pone.0226179>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

Analysis of genome-wide DNA arrays reveals the genomic population structure and diversity in autochthonous Greek goat breeds

S. Michailidou^{1,2*}, G. Th. Tsangaris³, A. Tzora⁴, I. Skoufos⁴, G. Banos^{1,5}, A. Argiriou², G. Arsenos¹

1 Laboratory of Animal Husbandry, School of Veterinary Medicine, School of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece, **2** Institute of Applied Biosciences, Center for Research and Technology Hellas, Thessaloniki, Greece, **3** Proteomics Research Unit, Biomedical Research Foundation of the Academy of Athens, Athens, Greece, **4** School of Agriculture, Department of Agriculture, Division of Animal Production, University of Ioannina, Kostakioi Artas, Greece, **5** Scotland's Rural College and The Roslin Institute University of Edinburgh, Edinburgh, Scotland, United Kingdom

* sofia_micha28@certh.gr



OPEN ACCESS

Citation: Michailidou S, Tsangaris GT., Tzora A, Skoufos I, Banos G, Argiriou A, et al. (2019) Analysis of genome-wide DNA arrays reveals the genomic population structure and diversity in autochthonous Greek goat breeds. PLoS ONE 14 (12): e0226179. <https://doi.org/10.1371/journal.pone.0226179>

Editor: Roberta Davoli, Università degli Studi di Bologna, ITALY

Received: October 23, 2018

Accepted: November 21, 2019

Published: December 12, 2019

Copyright: © 2019 Michailidou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Genotyping data (.ped and .map files) are publicly accessible via Zenodo database (<https://zenodo.org/record/3073175#XOPaAthRWht>).

Funding: This work was supported by the programs a) 'GOSHOMICS' under the action COOPERATION 2009 (SYNERGASIA 2009, project code 09SYN-23-990, grant awarded to GA), financed by the European Regional Development Fund and national resources, b) Data Driven

Abstract

Goats play an important role in the livestock sector in Greece. The national herd consists mainly of two indigenous breeds, the Eghoria and Skopelos. Here, we report the population structure and genomic profiles of these two native goat breeds using Illumina's Goat SNP50 BeadChip. Moreover, we present a panel of candidate markers acquired using different genetic models for breed discrimination. Quality control on the initial dataset resulted in 48,841 SNPs kept for downstream analysis. Principal component and admixture analyses were applied to assess population structure. The rate of inbreeding within breed was evaluated based on the distribution of runs of homozygosity in the genome and respective coefficients, the genomic relationship matrix, the patterns of linkage disequilibrium, and the historic effective population size. Results showed that both breeds exhibit high levels of genetic diversity. Level of inbreeding between the two breeds estimated by the Wright's fixation index F_{ST} was low ($F_{ST} = 0.04362$), indicating the existence of a weak genetic differentiation between them. In addition, grouping of farms according to their geographical locations was observed. This study presents for the first time a genome-based analysis on the genetic structure of the two indigenous Greek goat breeds and identifies markers that can be potentially exploited in future selective breeding programs for traceability purposes, targeted genetic improvement schemes and conservation strategies.

Introduction

The Greek national flock of goats is the largest in the European Union (E.U.), counting approximately 5 million heads [1]. These animals are classified into a wide range of farming systems; from semi-extensive low-input, traditional farms to large, semi-intensive, high producing and investing farms [2]. Most herds are traditionally reared in mountainous, semi-

Bioeconomy - DataBio (H2020 Grant Agreement No 732064, grant awarded to AA) and c) THRIVING LAND project, financed by the Trans Adriatic Pipeline (Grant Agreement HOR.044609, grant awarded to AA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

mountainous, lowland or insular areas, therefore are well adapted in different environmental conditions. For this reason, goat farming in Greece is preferred over other livestock systems in many marginal areas, because goats survive and produce in extreme environments and poor quality pastures (often including bushes and woody perennials), where other less versatile ruminants cannot.

Rearing of goats in Greece, traces back to Mycenaean times (about 1,200 B.C.); the first description on cheese making was reported by Homer and 500 years later, goats are found on silver coins in different Greek islands [3]. During the millennia of goat exploitation in Greece, many nuclei have evolved over time, forming different goat types, associated with their region of origin. Today, the vast majority of the Greek indigenous goats (90%) is comprised of the Eghoria breed, which is characterized by large variability in phenotypic characteristics. It is believed that Eghoria accounts for approximately 39 different types, which are mainly distinguished from their region of origin. The rest (10%), comprises of another autochthonous breed, the Skopelos (named after the homonymous island), which is a highly homogenous population of about 11,000 animals, as well as some foreign breeds (Damascus, Alpine and Saanen) and their crosses.

Herds of the Eghoria breed are found all over Greece, from isolated islands to inaccessible mountainous areas. This is a dual purpose breed, whose milk production ranges from 100 to 170 kg/lactation, depending on the rearing type. However, the genetic structure of this breed remains largely unknown, while it demonstrates large phenotypic, productive and reproductive variations. In contrast, Skopelos is a highly homogenous breed (based on phenotypic and productive traits), reared mainly at Northern Sporades island complex (Skopelos and Alonnisos islands). This breed has been intensively studied through national programs for the conservation of endangered indigenous breeds and for performance recordings, and has thus reached a milk yield of 225–260 kg/lactation. Historically, Skopelos and Eghoria breeds are assumed genetically distinct. The isolated geographical origin of Skopelos breed played a major role in supporting this premise. Yet, crossings between Eghoria and Skopelos breeds are usual, in order to increase milk performances in Eghoria and adaptability in more disadvantaged regions in Skopelos breed.

Unlike the cattle industry, where all Greek indigenous breeds are extinct or are in the verge of extinction due to the advent of more productive, high yielding animals (eg. Holstein-Friesian cattle), many indigenous breeds of small ruminants are still reared in Greece. However, during the last years, many goat herds are transforming, creating nuclei of crossbred animals (indigenous x imported purebreds). This diverse national herd has only been studied with few nuclear molecular markers, as single nucleotide polymorphisms (SNPs) [4, 5], amplified fragment length polymorphisms (AFLPs) [6] or mitochondrial DNA analysis [7]. During the last few years, the development of a genotyping microarray by the International Goat Genome Consortium (IGGC) [8], alongside with the availability of a goat reference genome [9, 10], have accelerated the understanding on domestication and genetic diversity in goat breeds, eliminating the errors introduced by the use of traditional genetic approaches.

Despite having the largest goat population in the E.U., Greece ranks in the 5th place in E.U. in total milk yield, after France, Czech Republic, Spain and Estonia, based on official data recorded for 2015 [1]. This is explained by the low productivity per doe, derived from the traditional and low input management systems applied in goat farming [2]. Although goat population in Greece is large compared to other livestock species, goat sector has a small impact in food industry in Greece. This reflects in analogous poor management systems and budgets, resulting in many cases in herds lacking rudimentary essentials like properly written and updated herd-books. However, intensive farming systems in Greece are more vulnerable, entailing higher levels of uncertainty due to increases in feed costs and changes in

environmental conditions [2]. Thus, in order to develop and render Greek goat farming sustainable, actions should be taken to accurately predict the genomic breeding value of an individual. Therefore, studying the genetic structure of a breed and understanding its genetic background will facilitate the genetic improvement in a population, as it is easier to monitor significant genes correlated with productivity, disease resistance or survival traits and therefore, easier to determine resilient individuals. To that end, the AdaptMap project (<http://www.goatadaptmap.org/>) is a valuable initiative led by international scientific groups that offers genotyping data of 130 breeds reared worldwide, so that can be used to monitor gene flow, migration events and genomic admixture levels with local or cosmopolitan breeds [11–16]. Such comparisons will help on the creation and preservation of purebred nuclei, which is of paramount importance as selected animals can be used in targeted breeding schemes utilizing the unique genetic resources.

Genomic characterization and protection of breeds has been extensively discussed in the past as a means to design breeding programs and future conservation strategies [17–19]. Moreover, breed traceability is a means to protect and valorize particular food products, and enhance consumer confidence towards foods of particular animal origin [20]. Yet, purebreds should be used with caution in selective breeding schemes, since inbreeding must be controlled to limit the potential impact of deleterious alleles and inbreeding depression on animal traits, and the loss of genetic variance [21, 22].

In the present study, we report a comprehensive genome-wide analysis of the population structure and genomic diversity of two Greek goat breeds, based on the genotypes generated by the Goat SNP50 BeadChip. Furthermore, we aim to identify genomic regions that could be used as potential markers to distinguish breeds and be further used in breed identification and conservation strategies.

Materials and methods

Ethics statement

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. In accordance to the 2010/63/EU guide and the adoption of the Law L276/33/20.10.2010 by the Greek Government (Law 106/vol A/30.04.2013), ethical approval is not required in our study. Blood sampling was conducted by veterinarians and/or under veterinarian supervision for routine veterinary care. All samples and data in our study were collected under the consent of the breeders.

Animal sampling

A total of 72 samples of two Greek domestic goat breeds (32 of Eghoria and 40 of Skopelos) were collected from 6 farms located in Central and Northern Greece. All animals were reared under semi-intensive conditions. Four milliliters of blood were collected from the jugular vein of each female goat in tubes containing EDTA as anticoagulant and stored in a freezer (-20°C) until further use in the laboratory. Breed discrimination was based on the available pedigree, morphology and productive trait records. Farm locations are indicated in a geographical map of Greece (S1 Fig) which was created in R [23] using the packages ‘maps’ [24], ‘rworldmap’ [25] and ‘rwordextra’ [26]. Details on farms locations and number of animals sampled per farm and breed are presented on S1 Table.

DNA extraction and genotyping

DNA was extracted from blood samples using the Nucleospin Blood Quick Pure kit (Macherey-Nagel, Germany) according to the manufacturer’s protocol. DNA integrity was

assessed by electrophoresis on a 0.8% TAE agarose gel. DNA concentration was quantified with the Qubit® dsDNA BR assay kit (Qubit 2.0 Fluorometer, Invitrogen, Carlsbad, CA, USA). All animals were genotyped using Illumina's Goat SNP50 BeadChip containing 53,347 SNPs. DNA samples were dried down in 1 mM Tris-EDTA and 400 ng of each sample was sent to GeneSeek (Neogen Corporation, UK) for genotyping on an Illumina's iScan platform. Imaging and data generation were performed with the Illumina's iScan system.

Data analysis

Sample and SNP quality control. Genotypes were generated and pre-processed by the iScan system based on an automated genotype calling feature by Illumina's GenomeStudio software (v.1.9.4). During this step a preliminary quality control on individuals (exclusion or inclusion of samples) and basic statistics for each SNP are generated e.g. call rate, allele frequencies. Filtering of samples was performed based on the call rate for each sample (call rate > 0.9) generated by Illumina's GenomeStudio software, as well as using the *mind* command in PLINK v1.90 [27]. Chromosomal coordinates for each SNP were mapped to the updated goat reference ARS1 assembly [9].

PLINK v1.90 was used to generate summary statistics and conduct exact tests for deviation from Hardy-Weinberg equilibrium (HWE). Quality control and SNP filtering were performed using PLINK v1.90 [27] by filtering out SNPs with minor allele frequency (MAF) < 1%, call rate < 0.98, HWE p-value $\leq 1.0E-6$, those lacking genomic location and sex-linked SNPs, since they exhibit lower mutation rates and a direct comparison of the overall diversity between the X chromosome and autosomes is difficult [28, 29].

Population structure analysis. Assessment of the genetic structure of individuals was performed in autosomal filtered SNPs using the ADMIXTURE v1.3 software [30], assuming a number of subpopulations (K) ranging from 2 to 5 and visualized using R plots. The optimal number of K value was determined as the one having the lowest cross-validation (CV) error. Population structure was also examined by principal component analysis (PCA) in autosomal SNPs; PLINK was used for the generation of eigenvectors and eigenvalues which were visualized with the GENESIS software [31] demonstrating the relationship between PC1 and PC2.

Genetic structure of the two Greek goat breeds was also analyzed including data from 45 breeds reared worldwide, in order to assess whether admixture with cosmopolitan breeds has occurred in Greek breeds. Genotypes were obtained from the AdaptMap project [12] and are available in <https://doi.org/10.5061/dryad.v8g21pt>. Selection of the 45 goat breeds was based on the transit of goats from their putative center of domestication to Europe, both from the Danubian corridor and the Mediterranean basin. Thus, data from 29 breeds from Europe, 3 breeds from Near East, 9 breeds from Northern Africa and 4 breeds from eastern Asia were selected for comparison purposes with Greek data. Quality control and SNP filtering were also performed using PLINK v1.90, by filtering out SNPs in autosomes with minor allele frequency (MAF) < 1% and call rate < 0.98. Population substructure among the 47 goat populations using ADMIXTURE software was conducted assuming a number of K from 2 to 50. Visualizations of ancestry coefficients estimated with ADMIXTURE software were performed using the BITE R package [32].

To further investigate historical gene flow between the populations analyzed, we used Tree-mix software v.1.12 [33]. Maximum likelihood trees were constructed for different migration events, starting from *m0* (no migration event) to *m15*, by grouping SNPs in windows of 200. To increase robustness of analysis, we applied three replicates per migration event. The percentage of fraction of variance explained for each *m*, the log-likelihood value for each *m* and the residuals from the fit of the model among individuals were used to determine and visualize

the most predictive model (m_{best}). Then, nodes robustness was estimated for the m_{best} by running 100 bootstrap replicates using the *Treemix_bootstrap.sh* script and plotted using the *treemix.bootstrap* R function, both implemented in BITE R package [32].

Genetic diversity indices and inbreeding levels. Observed (H_o) and expected (H_e) heterozygosities were calculated for each breed separately to measure the genetic diversity within breed using ARLEQUIN 3.5.2.2 software [34], by applying a correction on the number of usable SNPs proposed by Colli et al. [14]. Wright's inbreeding coefficient F_{IS} (Individual within Subpopulation) was calculated per animal using PLINK. Genomic inbreeding was also calculated following Van Raden [35]; computation of inbreeding values were assessed from the diagonal of the genomic relationship matrix, denoted as F_{GRM} . Estimates of inbreeding coefficients were calculated from autosomes, for each animal separately, using the GCTA program [36] starting from the PLINK binary PED files (.bed, .fam, .bim). Wright's pairwise F_{ST} value was calculated using ARLEQUIN software to test inbreeding levels and genetic distance between breeds.

Runs of homozygosity. Runs of homozygosity (ROH) were defined in all autosomes to assess inbreeding levels for all animals, in each breed separately, and categorized based on their length and chromosome using PLINK v1.90. ROHs were calculated with the 'Runs of homozygosity' function in PLINK software with adjusted parameters for the total length of ROHs (≥ 1 Mb), the variants in the scanning window ($n = 15$), the number of heterozygous SNP allowed ($n = 1$) and the number of missing calls ($n = 1$) to estimate homozygosity. The percentage of ROHs per chromosome was calculated as proposed by Al-Mamun et al. [37], as follows:

$$\text{Average percentage of ROH per chromosome} = \frac{\sum \text{ROHs in Mbp per chromosome}}{N * \text{Chromosome length (Mbp)}} * 100$$

where N is the number of animals that had a ROH in that chromosome.

Genomic inbreeding coefficient based on ROHs (F_{ROH}) was calculated for each animal from the sum of ROH lengths, divided by the total length of the autosomal genome (kb) covered by SNPs as proposed by McQuillan et al. [38]. Inbreeding coefficients based on ROHs were calculated for four bins, grouped according to the length of ROHs i) all ROHs (F_{ROH}), ii) < 10 Mb ($F_{ROH < 10Mb}$), iii) 10–20 Mb ($F_{ROH 10-20Mb}$) and iv) > 20 Mb ($F_{ROH > 20Mb}$). In all cases, coefficients were calculated separately for each animal and then averaged within breed.

ROH frequencies were calculated for each breed and for six different length categories (< 3 Mbp, 3–5 Mbp, 5–10 Mbp, 10–20 Mbp, 20–30 Mbp, > 30 Mbp), using the same classification reported by the AdaptMap project for comparison purposes [12]. ROH frequencies were plotted for each length category and for each breed separately, by summarizing all ROHs and averaged against the total length of ROHs (in Mbp).

To identify SNPs that are present inside ROHs the R package 'detectRUNS' was used [39]. The genome-wide occurrence of SNPs in ROHs was expressed as the proportion (%) of times each SNP fall inside the defined ROHs, plotted against SNP position per chromosome, for each breed separately.

Linkage disequilibrium analysis. Linkage disequilibrium (LD) was tested for each breed separately to examine recombination of linked SNPs using PLINK v1.90 with default parameters; SNPs included in this step spanned a distance from 0.001 to 1 Mb. The squared correlation coefficient (r^2) curve was estimated by determining the nonlinear least squares fit line using the *nls* function in R. The r^2 coefficient was used instead of D' as a more reliable measurement in studies with small sample sizes and more useful in predicting the power of association mapping [40].

Effective population size analysis. Effective population size (N_e) was estimated separately for each breed, using SNeP v.1.1 [41]. N_e estimates at different generations were based on linkage disequilibrium using the formula suggested by Corbin and co-authors [42]. Estimated effective population size was plotted over the last 150 and 1,000 generations to assess relevant diversity trends.

Discriminatory SNPs among breeds. In order to identify regions putatively under selection and SNPs that can be used for the discrimination of breeds, three different approaches were carried out:

- a. Calculation of F_{st} for each marker: F_{st} value for each SNP was calculated with the `fst-within` command in PLINK v1.90, using the method introduced by Weir and Cockerham (W&C) [43]. SNPs at a threshold corresponding to the 0.995 percentile of the total distribution were acquired for gene annotation. Manhattan plots demonstrating the F_{st} value for each SNP were constructed using the qqman R package [44].
- b. Identification of discriminatory SNPs for breed assignment using the Toolbox for Ranking and Evaluation of SNPs (TRES) software [45]. Evaluation of SNPs was performed by comparing SNPs obtained from all available methods; Delta [46], Wright's pairwise F_{st} [47], and Informativeness for Assignment [48]. From each analysis, the top 200 SNPs were required, and only common SNPs among methodologies were further evaluated. Using the TRES software two methodologies were followed:
 - i. Identification of discriminatory SNPs was performed using the whole dataset of animals per breed, denoted as "TRES_all".
 - ii. Identification of discriminatory SNPs was performed by randomly splitting the dataset into training and test populations (using the `awk srand(n)` function), denoted as "TRES_tt".

Consensus SNPs among the three methods (F_{st} , TRES_all, TRES_tt) were obtained and evaluated using GeneClass2 [49]. SNPs per methodology were submitted to Venny 2.1 [50] to depict common and/or shared SNPs. For the evaluation of discriminatory SNPs, two approaches were followed; assignment or exclusion of individuals based on the discriminatory SNPs and detection of first generation migrants using the likelihood (L) estimation calculated from L_{home}/L_{max} [51]. In both approaches, SNPs were evaluated using frequency-based [52] and Bayesian [53] criteria. In all cases, Monte Carlo resampling was enabled, using the simulation algorithm proposed by Paetkau et al. [51] with a number of simulated individuals of 1,000 and a type I error (alpha) threshold of 0,001.

Discriminatory SNPs obtained from all methods were annotated on the ARS1 assembly [9] using the Genome Data Viewer from National Center for Biotechnology Information (NCBI). Annotation was performed to reveal genes or nearby genes (within $\pm 100\text{kb}$) from the positions of identified SNPs that might indicate signatures of selection for the two breeds.

Results

Sample and SNP filtration, basic descriptive statistics

No sample was excluded from further analysis; mean call rate for the 72 samples was 0.974. Quality control of SNPs was performed by filtering out non-informative SNPs for minor allele frequency (MAF), call rate (call frequency) and HWE. As such, 4,506 SNPs were excluded, resulting in 48,841 SNPs kept for downstream analysis (S2 Table). Genotyping data (.ped and .map files) are publicly accessible via Zenodo database (<https://zenodo.org/record/3073175#.XOPaAthRWHt>).

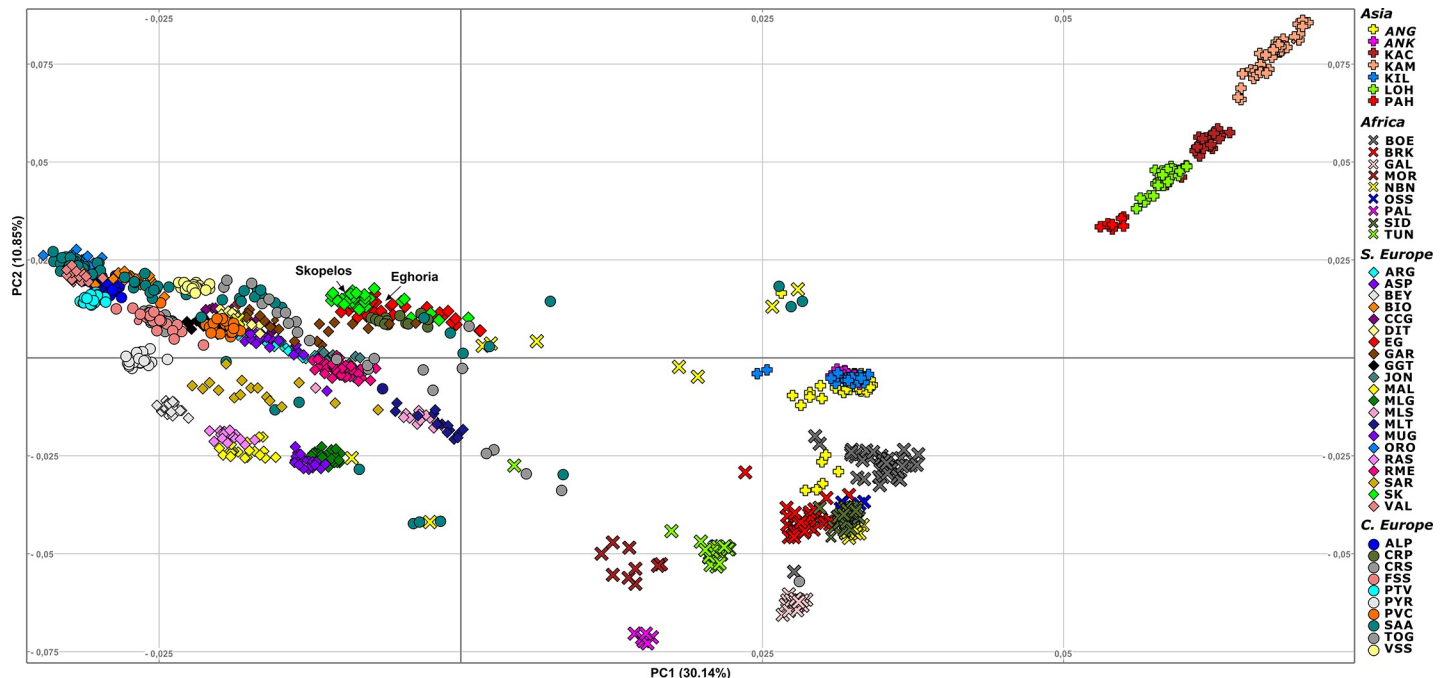


Fig 1. Principal component analysis of the first two axes in 1,212 goat samples from 47 breeds. EG: Eghoria; SK: Skopelos; ALP: Alpine; ANG: Angora; ANK: Ankara; ARG: Argentata; ASP: Aspromontana; BEY: Bermeya; BIO: Bionda dell'Adamello; BOE: Boer; BRK: Barki; CCG: Ciociara Grigia; CRP: Carpatian; CRS: Corse; DIT: Di Teramo; FSS: Fosses; GAL: Galla; GAR: Garganica; GGT: Girgentana; JON: Jonica; KAC: Kachan; KAM: Kamori; KIL: Kil; LOH: Lohri; MAL: Mallorquina; MLG: Malaguenia; MLS: Maltese Sarda; MLT: Maltese; MOR: Moroccan goat; MUG: Murciano-Granadina; NBN: Nubian; ORO: Orobica; OSS: Oasis; PAH: Pahari; PAL: Palmera; PTV: Poitevine; PVC: Provencale; PYR: Pyrenean; RAS: Blanca de Rasquera; RME: Rossa Mediterranea; SAA: Saanen; SAR: Sarda; SID: Saidi; TOG: Toggenburg; TUN: Tunisian; VAL: Valdostana; VSS: Valpassiria.

<https://doi.org/10.1371/journal.pone.0226179.g001>

Population structure analysis

The genetic structure among the two Greek breeds and the 45 selected breeds reared world-wide, assessed from PCA, revealed three major clusters; one enclosing European breeds, a second enclosing only breeds from Pakistan (Central Asia) and a third looser cluster consisting of breeds from Asia and Africa (Fig 1). Results showed that Greek breeds cluster together with European breeds and are found in greater distance from the other two clusters. From the breeds that clustered near Greek breeds, PCA revealed that Eghoria and Skopelos grouped closer to the Carpatian breed (CRP), as well as to Italian breeds, such as Garganica (GAR), Rossa Mediterranea (RME) and Jonica (JON). Admixture analysis revealed that Greek breeds maintain a distinct genetic profile compared to other European breeds, although cosmopolitan breeds like Alpine and Saanen are reared from many decades in Greece (Fig 2). The lowest CV error for the 47 breeds was acquired for $K = 37$.

Treemix analysis verified ADMIXTURE and PCA results, since no gene flow between Alpine and Saanen and Greek breeds is observed at $m10$ (Fig 3). Overall, the fraction of variance explained in our dataset ranged from 89.49% ($m0$) to 95.03% ($m15$). The model with the 10 migration edges was chosen as the most predictive model since it explains 93.52% of the variance, and all migration events thereafter ($m11$ to $m15$) did not account for much more variance (S2 Fig). Moreover, identification of the populations that are not well-modeled at $m10$ presented only a few pairs of populations, mostly of Italian origin, with high standard error that are not well explained, thus are candidates for admixture events (S3 Fig). The consensus tree obtained from the 100 replicates for the Greek breeds showed that their nodes were supported by bootstrap values below 50. Eghoria and Skopelos breeds were located in separate

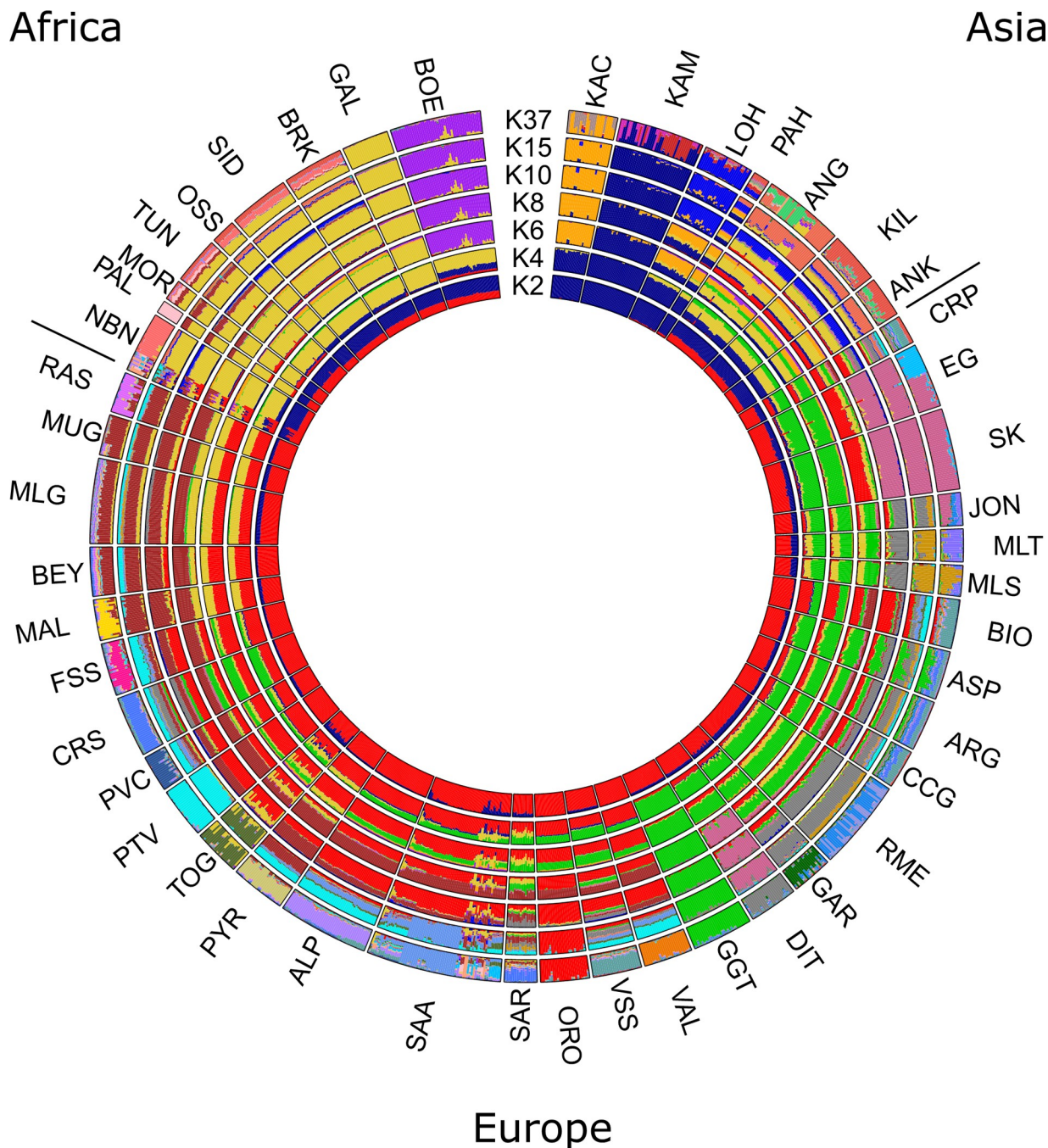


Fig 2. Circular representation of Admixture analysis at $K = 2, 4, 6, 8, 10, 15$ and 37 for $1,212$ goat samples from 47 breeds. Each individual is presented by a vertical bar. Different colors indicate different clustering groups. EG: Eghoria; SK: Skopelos; ALP: Alpine; ANG: Angora; ANK: Ankara; ARG: Argentata; ASP: Aspromontana; BEY: Bermeya; BIO: Bionda dell'Adamello; BOE: Boer; BRK: Barki; CCG: Ciociara Grigia; CRP: Carpatian; CRS: Corse; DIT: Di Teramo; FSS: Fosses; GAL: Galla; GAR: Garganica; GGT: Girgentana; JON: Jonica; KAC: Kachan; KAM: Kamori; KIL: Kil; LOH: Lohri; MAL: Mallorquina; MLG: Malaguena; MLS: Maltese Sarda; MLT: Maltese; MOR: Moroccan goat; MUG: Murciano-Granadina; NBN: Nubian; ORO: Orobica; OSS: Oasis; PAH: Pahari; PAL: Palmera; PTV: Poitevine; PVC: Provencale; PYR: Pyrenean; RAS: Blanca de Rasquera; RME: Rossa Mediterranea; SAA: Saanen; SAR: Sarda; SID: Saidi; TOG: Toggenburg; TUN: Tunisian; VAL: Valdostana; VSS: Valpassiria.

<https://doi.org/10.1371/journal.pone.0226179.g002>

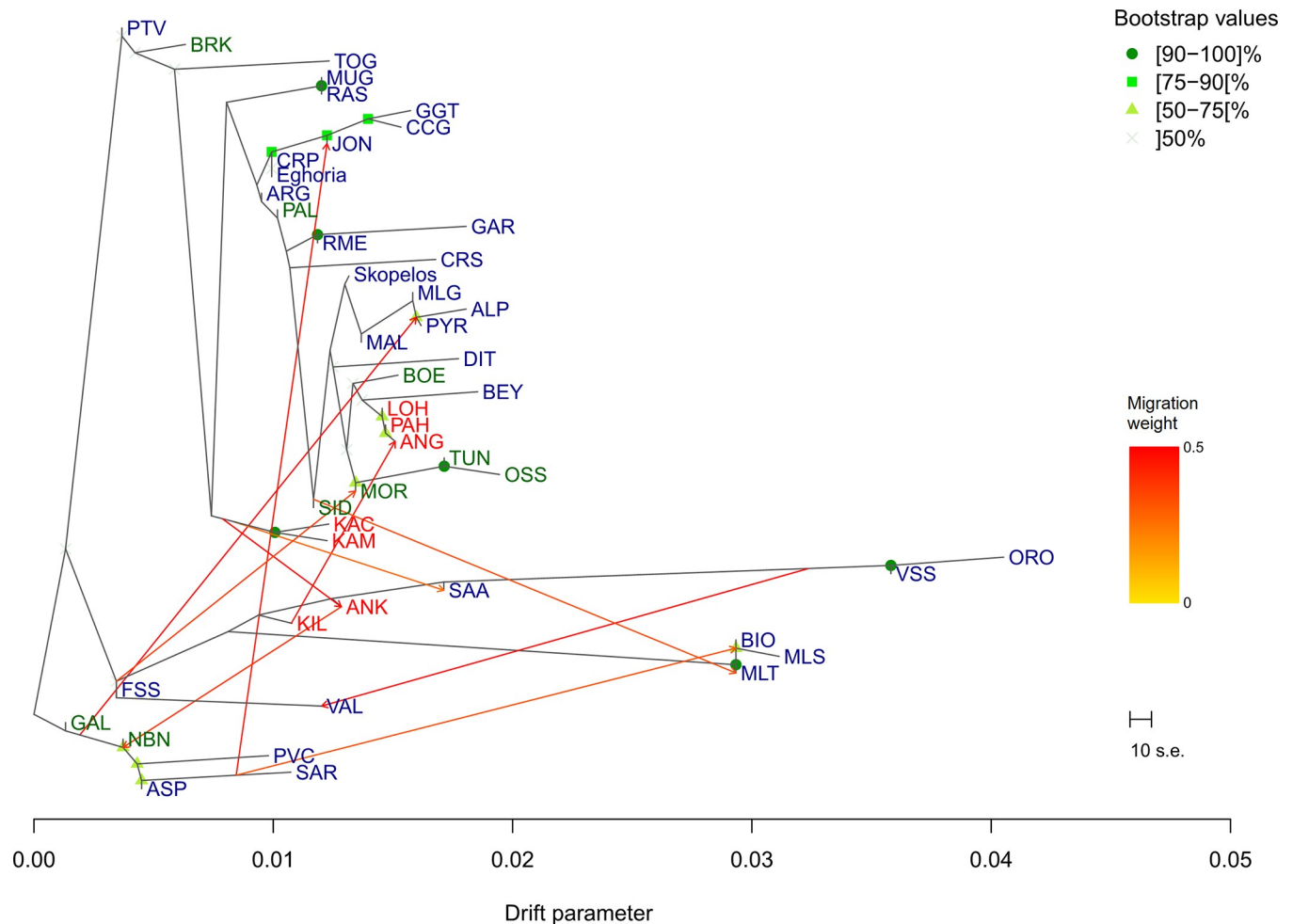


Fig 3. Treemix analysis with 10 migration events. Nodes robustness was estimated with 100 bootstrap replicates. Bootstrap values below 50 are not shown. Migration edges are colored according to their migration weight. Breeds are colored according to their geographical origin (blue for European, red for Asian and green for African breeds). ALP: Alpine; ANG: Angora; ANK: Ankara; ARG: Argentata; ASP: Aspromontana; BEY: Bermeya; BIO: Bionda dell'Adamello; BOE: Boer; BRK: Barki; CCG: Ciociara Grigia; CRP: Carpatian; CRS: Corse; DIT: Di Teramo; FSS: Fosses; GAL: Galla; GAR: Garganica; GGT: Girgentana; JON: Jonica; KAC: Kachan; KAM: Kamori; KIL: Kil; LOH: Lohri; MAL: Mallorquina; MLG: Malaguena; MLS: Maltese Sarda; MLT: Maltese; MOR: Moroccan goat; MUG: Murciano-Granadina; NBN: Nubian; ORO: Orobica; OSS: Oasis; PAH: Pahari; PAL: Palmera; PTV: Poitevine; PVC: Provencale; PYR: Pyrenean; RAS: Blanca de Rasquera; RME: Rossa Mediterranea; SAA: Saanen; SAR: Sarda; SID: Saidi; TOG: Toggenburg; TUN: Tunisian; VAL: Valdostana; VSS: Valpassiria.

<https://doi.org/10.1371/journal.pone.0226179.g003>

nodes, grouped together however with Europeans breeds. No major migration event was observed for the Greek breeds at $m10$. Skopelos breed was found as a surrounding population (together with MLG, MAL and ALP) to a high-weighted edge that links Galla breed (GAL) from Kenya to Pyrenean breed (PYR). Eghoria breed was located in a node enclosing Carpatian (CRP), Jonica (JON), Girgentana (GGT) and Ciociara Grigia (CCG) breeds. This clade was found to be linked with a high-weighted edge originating from Italian breeds Aspromontana (ASP) and Sarda (SAR).

Population substructure was also assessed exclusively in Greek goats, using both ADMIXTURE and PCA analysis in autosomal filtered SNPs in order to evaluate purebreds and detect possible crossbreds. Population structure was investigated assuming a number of K from 2 to 5. Cross validation error was the lowest for $K = 3$ (0.63661), indicating the most likely number of different sub-populations represented in the 72 samples [30]. At $K = 2$ the first group that

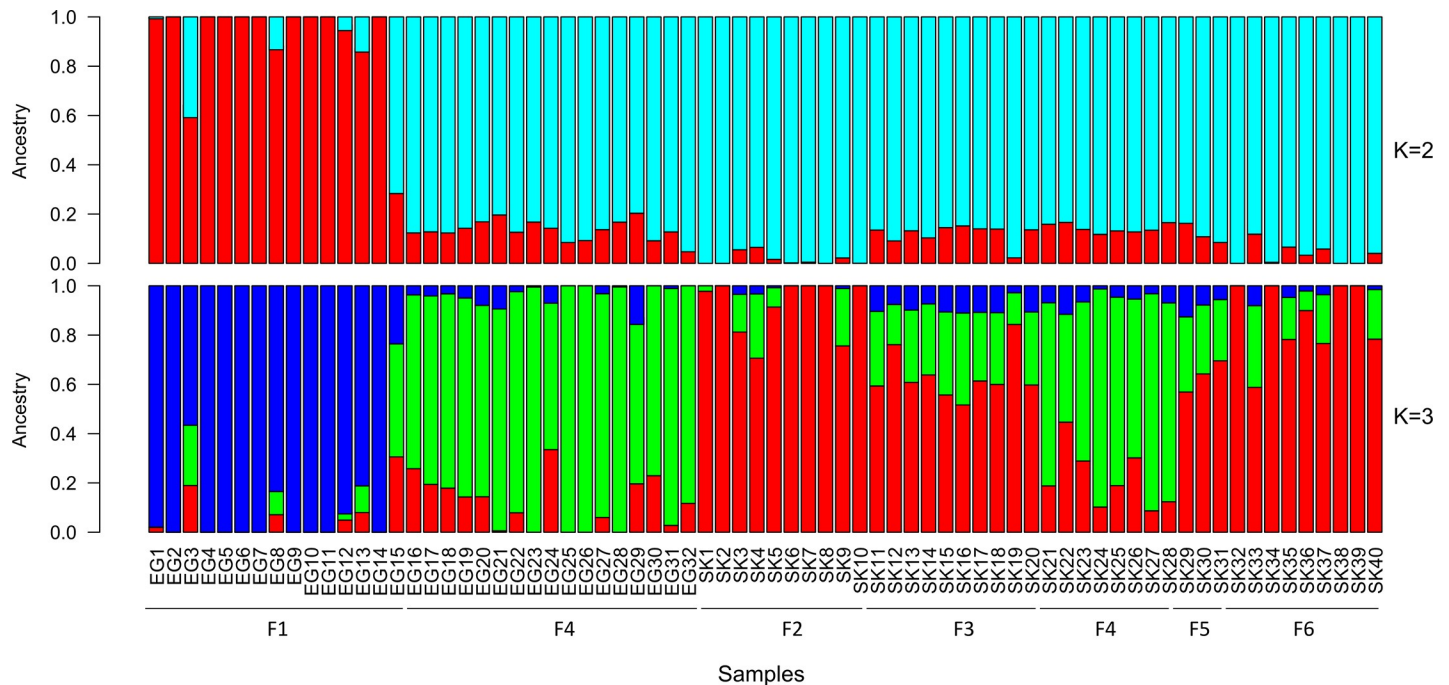


Fig 4. Admixture analysis at K = 2 and 3 for the 48,841 autosomal SNPs in Greek breeds. Each individual is represented by a vertical bar. Different colors indicate different clustering groups. F1: farm 1; F2: farm 2; F3: farm 3; F4: farm 4; F5: farm 5; F6: farm 6. SNPs: Single nucleotide polymorphisms.

<https://doi.org/10.1371/journal.pone.0226179.g004>

differentiated consisted of individuals of the Eghoria breed (samples EG1 to EG14), all originating from the same farm (farm 1 in S1 Fig) located in a geographical isolated region in the Pindus mountain range (Fig 4). At K = 3, this nucleus remained differentiated from the others and two additional clusters were formed. Individuals of Skopelos breed exhibited a more uniform profile, including however eight samples (SK21 to SK28) originating from the municipality of Thessaloniki (farm 4 in S1 Fig) that had a somewhat different genetic profile. Notably, this particular nucleus of Skopelos individuals had the same genetic profile with many samples of Eghoria breed (samples EG16 to EG32), originating from the same farm, too. At K = 4 the clusters that were formed were similar to those obtained at K = 3. The main difference was that the classification of individuals of farm 1 started to disappear progressively. At the highest K value of 5, cross-validation error reached 0.67373 and a weaker population substructure was observed, maintaining however the main trend in the clustering profile. From the aspect of farms, the mean component values per farm and K were calculated to reveal potential geographic differentiation patterns; similar clustering profiles were acquired with farm 1 being the most differentiated, followed by farm 4 (S4 Fig).

PCA analysis in the 48,841 autosomal SNPs revealed the same clustering profile with ADMIXTURE analysis. The first two principal components (PC1 and PC2) explained 26.02% of the total genetic variation. In particular, three distinct clusters were formed, associated with region of origin (Fig 5). Analysis of principal components showed that Eghoria breed possesses higher levels of genetic variation, compared to Skopelos breed. Individuals of the Eghoria breed originating from farm 1 formed a distinct cluster in a greater distance from all the remaining samples, which was also the case in clustering at K = 2. From the other two clusters, the one consisted only of Skopelos individuals who are reared in their region of origin (farms 2, 3, 5, 6 in S1 Fig) and the third, the one with the admixed population, consisted of samples from both breeds, which were reared in the same farm. Considering that the admixed animals

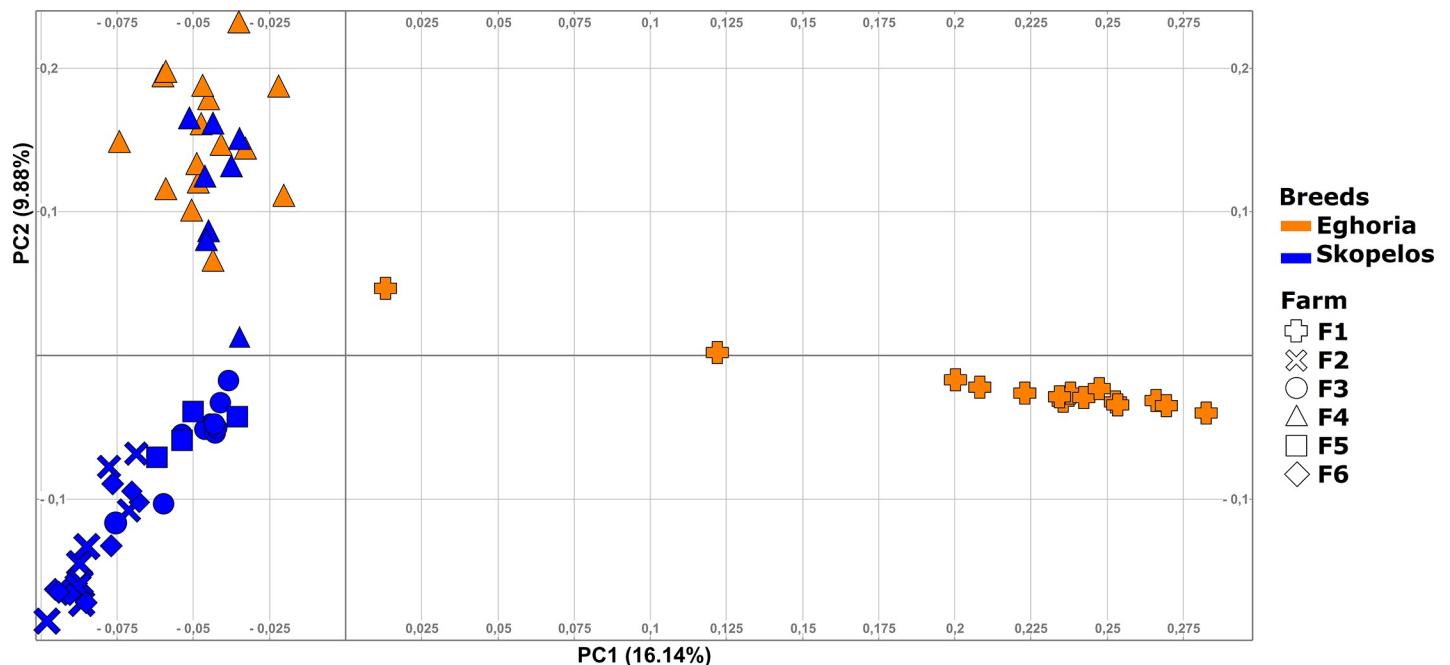


Fig 5. Principal component analysis of the first two axes in 72 goat samples using the 48,841 SNPs. SNPs: Single nucleotide polymorphisms.

<https://doi.org/10.1371/journal.pone.0226179.g005>

of Skopelos breed could introduce errors in the estimation of genetic heterozygosity and inbreeding indices, they were excluded from downstream analysis.

Within and between breed genetic diversity

After the exclusion of admixed animals, the new dataset consisted of 64 animals in total (32 female goats for each breed). From the 48,841 quality-filtered SNPs, the percentage of within-breed polymorphic loci was greater than 95% for both breeds (Table 1). The number of polymorphic loci was 46,608 and 46,732 SNPs for Eghoria and Skopelos breeds, respectively. Even though only the Skopelos breed had been used for the validation of Goat SNP50 BeadChip [8], the Eghoria breed also maintained high levels of polymorphic loci. According to H_o and H_e values, Eghoria breed presented higher within-breed level of genetic diversity than Skopelos, yet differences between breeds were not substantial. Mean H_e values were 0.405 ± 0.110 and 0.394 ± 0.120 for Eghoria and Skopelos breeds, respectively. Intra-population nucleotide diversity estimated from π values, was quite similar between breeds (0.404 and 0.392 for Eghoria and Skopelos breeds, respectively).

Different estimates of inbreeding coefficients were calculated from the SNP genotyping data. Wright's inbreeding coefficient F_{IS} was calculated separately for each individual; both breeds presented positive mean F_{IS} values (0.0312 and 0.0376 for the Eghoria and Skopelos breeds, respectively) indicating more homozygotes in the studied population than expected

Table 1. Genetic diversity within breeds derived from individuals per breed (N), as measured by the number (N_p) and percentage of polymorphic loci ($N_{p\%}$), observed (H_o) and expected (H_e) heterozygosity with standard deviation ($\pm sd$) and nucleotide diversity per breed (π), for the 48,841 filtered single nucleotide polymorphisms (SNPs).

Breed	N	N_p	$N_{p\%}$	$H_o (\pm sd)$	$H_e (\pm sd)$	π
Eghoria	32	46,608	95.42%	0.395 ± 0.132	0.405 ± 0.110	0.404
Skopelos	32	46,732	95.68%	0.392 ± 0.140	0.394 ± 0.120	0.392

<https://doi.org/10.1371/journal.pone.0226179.t001>

Table 2. Average inbreeding coefficients and standard errors (SE) per breed, estimated by Wright's inbreeding coefficient (F_{IS}), genomic relationship matrices inbreeding coefficient (F_{GRM}) and derived from runs of homozygosity inbreeding coefficient (F_{ROH}) for different length categories.

Breed	F_{IS} (\pm SE)	F_{GRM} (\pm SE)	F_{ROH} (\pm SE)	F_{ROH} (\pm SE)		
				$F_{ROH < 10Mb}$ (\pm SE)	$F_{ROH 10-20Mb}$ (\pm SE)	$F_{ROH > 20Mb}$ (\pm SE)
Eghoria	0.0312 (\pm 0.0151)	0.0452 (\pm 0.0146)	0.0680 (\pm 0.0112)	0.0009 (\pm 0.0000)	0.0000 (\pm 0.0000)	0.0712 (\pm 0.0112)
Skopelos	0.0376 (\pm 0.0062)	0.0211 (\pm 0.0048)	0.0287 (\pm 0.0037)	0.0038 (\pm 0.0000)	0.0060 (\pm 0.0005)	0.0330 (\pm 0.0038)

<https://doi.org/10.1371/journal.pone.0226179.t002>

(Table 2). Lower inbreeding coefficient values were obtained for the Skopelos breed compared to Eghoria, in the other two indices studied (F_{GRM} and F_{ROH}) (Table 2). The mean value of the genomic estimator F_{GRM} was positive for both breeds, but very close to 0 (0.0452 for Eghoria and 0.0211 for Skopelos breeds), indicating that the variance is low. Inbreeding coefficient based on all ROHs revealed higher levels of autozygosity in Eghoria ($F_{ROH} = 0.0680$) than in Skopelos breed ($F_{ROH} = 0.0287$). Among the different F_{ROH} bins, $F_{ROH > 20Mb}$ expressed the highest values for both breeds. Moreover, at small lengths, where very short and common ROHs are located due to LD [54], $F_{ROH < 10Mb}$ presented very low values in both breeds. Overall, differences on levels of inbreeding reflected by F_{IS} , F_{ROH} and F_{GRM} were quite modest; similar levels of inbreeding were found between the two breeds. Genetic differentiation of breeds and the level of their relatedness, indicated by pairwise F_{ST} values was low (0.04362).

Runs of homozygosity

According to the parameters used, 765 ROHs were found in total; Eghoria had a slightly larger number of ROHs ($n = 389$) than Skopelos ($n = 376$) with an average of 12.16 and 11.75 ROHs per animal, respectively, including those that did not present any ROH (S5 Fig). The longest ROH was found in Eghoria breed (max. length in sample EG6 = 69.336 Mbp) which did not differ significantly in length compared to the longest ROH in Skopelos (max. length 60.261 Mbp for sample SK6). On average, Eghoria presented largest mean length of ROHs for the entire population (mean length = 9.488 Mbp), compared to Skopelos breed (mean length = 6.022 Mbp). Among the 765 ROHs, 51 ROHs were longer than 20 Mbp (39 for Eghoria and 12 for Skopelos), 115 ROHs were between 10 to 20 Mbp (74 for Eghoria and 41 for Skopelos) and 599 ROHs were found below the length of 10 Mb (276 for Eghoria and 323 for Skopelos breeds). Nevertheless, Eghoria demonstrated considerable differences compared to Skopelos breed concerning the total length of genome covered by ROHs (Fig 6). For instance, EG6 expresses 34 ROH segments covering over 500 Mbp of its genome, whereas, SK6 counts 27 ROH segments of 214.73 Mbp in total.

ROHs frequencies were further classified according to their size and breed. For both breeds, the frequency of short ROHs was higher compared to ROHs of greater size (Fig 7). In general, results showed that there is an inverse trend between ROH frequency and ROH length, indicating the absence of recent inbreeding among individuals within each breed. Again, no significant differences were observed among breeds.

Chromosome 6 had the largest number of ROHs ($n = 48$), followed by chromosome 10 ($n = 47$) and chromosome 1 ($n = 46$). Overall, the total number of ROHs per chromosome decreased with decreasing chromosome length (S6 Fig), except for chromosomes 2, 3, 4 and 5 that presented low number of ROHs and chromosome 21 that presented high number of ROHs, compared to their size. Looking at the percentage of ROHs per chromosome, the highest percentage was found in chromosome 26 (24.17%) followed by chromosome 25 (24.09%) and the lowest was found in chromosome 5 (6.06%). The average percentage of ROHs per chromosome showed an inverse relationship, with these values increasing while chromosome size decreases. Regarding SNPs located within ROHs and across autosomes, chromosome 6

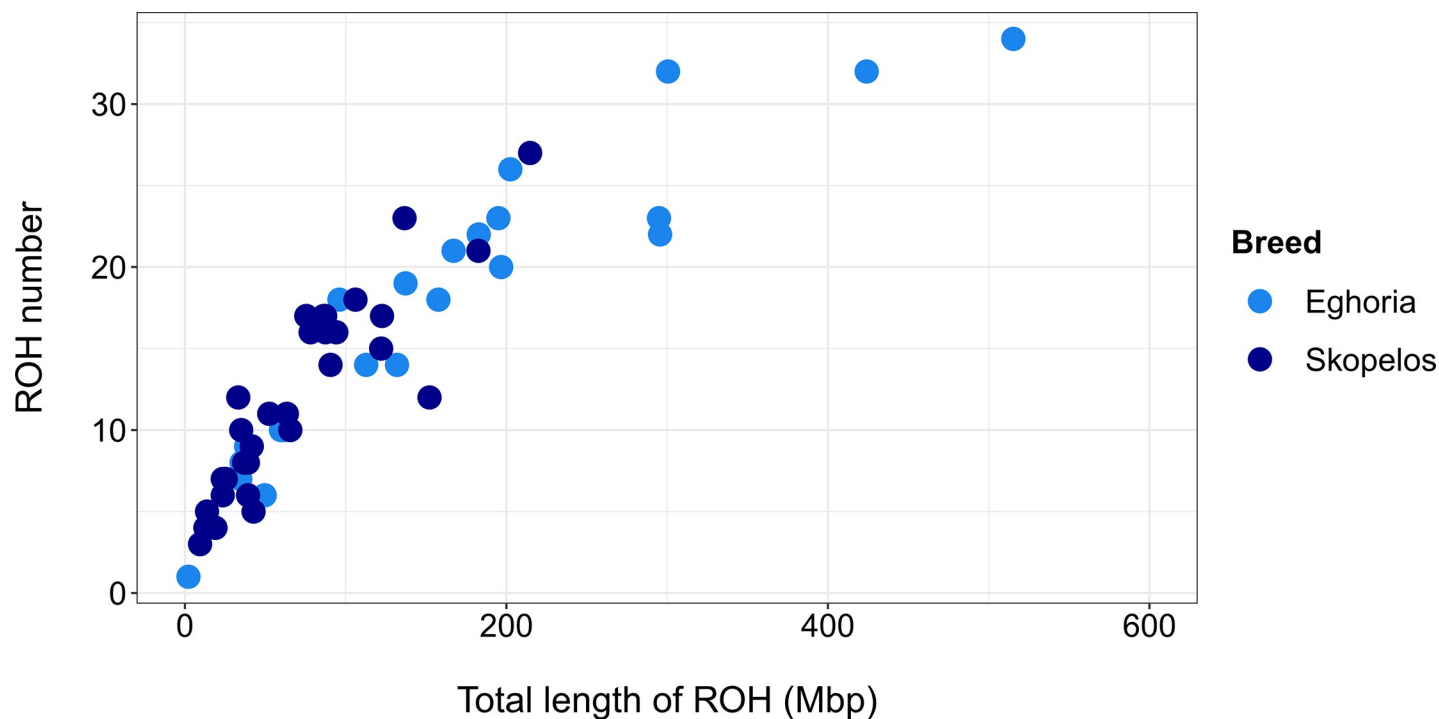


Fig 6. Total number of runs of homozygosity (ROHs) per animal and breed, compared to the total length of ROHs.

<https://doi.org/10.1371/journal.pone.0226179.g006>

had the highest frequency of SNPs in ROH segments for both breeds (S7 Fig). Although the proportion of SNPs in ROHs is below 30% for both breeds and on any chromosome, homozygous clusters can still be spotted throughout the genome.

Linkage disequilibrium patterns

Extent of linkage disequilibrium was assessed with pairwise r^2 in the 48,841 autosomal SNPs, for each breed separately. The pattern of LD measured by r^2 was quite similar between breeds.

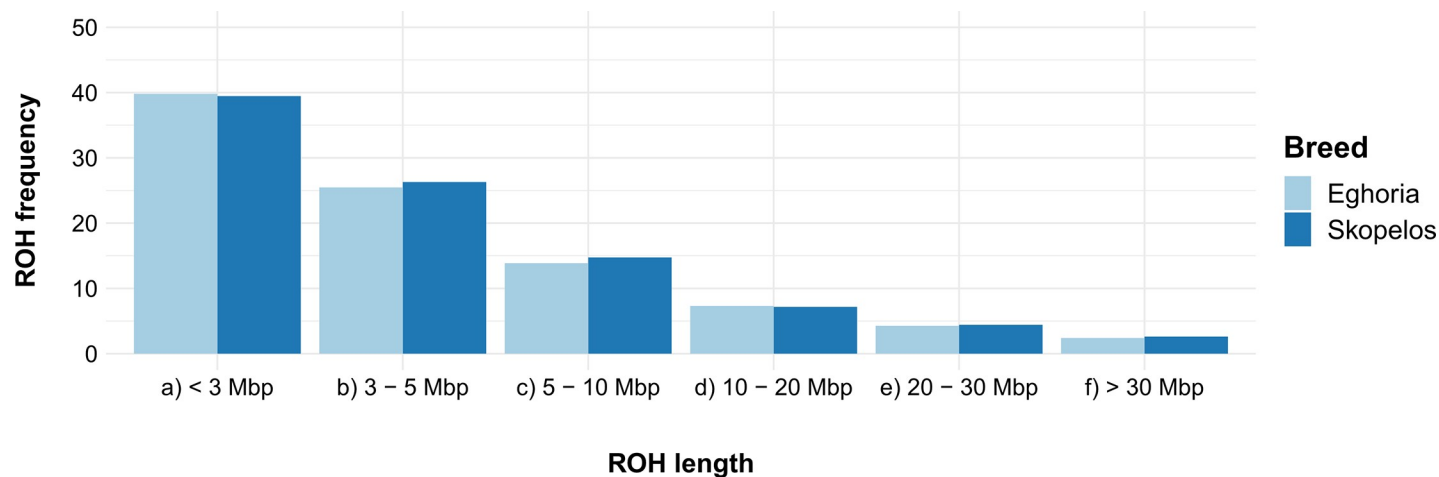


Fig 7. Runs of homozygosity (ROHs) frequencies per breed, classified according to their length. ROHs were classified in six length categories (<3 Mbp, 3–5 Mbp, 5–10 Mbp, 10–20 Mbp, 20–30 Mbp, >30 Mbp). Mbp: Megabase pair.

<https://doi.org/10.1371/journal.pone.0226179.g007>

The most rapid decline for both breeds was observed over the first 0.1 Mbp. Eghoria breed showed somewhat higher rates of LD decay compared to Skopelos, which displayed slightly increased levels of LD (S8 Fig). In general, LD patterns measured by r^2 were alike between breeds according to the parameters used; mean r^2 was 0.084 in Eghoria and 0.087 in Skopelos breed. The average inter-marker distances were about 260 kb for both breeds. Chromosomes 1 and 2 had the largest number of adjacent SNPs in LD, whereas the lowest was observed in chromosome 25; generally, the total number of adjacent SNPs in LD tend to decrease with decreasing chromosome length.

Effective population size

Estimates of ancestral N_e obtained for 26 time points, for the past ~1,000 generations are presented in S3 Table. Effective population size in both breeds displayed a decreasing trend over time. According to the genotyped populations, in the distant past (more than 121 generations ago) Eghoria had larger N_e compared to Skopelos breed. However, in recent past (13–98 generations ago), Eghoria presented lower N_e values compared to Skopelos. In the distant past, for the period of ~1,000 generations ago, N_e was estimated to be 3,659 and 3,391 for the Eghoria and Skopelos breed, respectively (Fig 8A). The most recent N_e values dated 13 generations ago, which equals to ~52 years ago assuming a generation interval of 4 years in goats, with the respective numbers being reduced to 96 and 127, representing a narrower genetic pool for both breeds (Fig 8B).

Discriminatory SNPs between breeds

Genome-wide association analysis was applied to the dataset of 48,841 SNPs for all animals, to identify markers for breed discrimination and origin assignment. To that end, three approaches were used. Using W&C's F_{st} approach, 151 SNPs were detected at a threshold of 0.372 (equal to the 0.995 of the percentile distribution) (Fig 9). Discriminatory SNPs were located on 28 autosomes; no SNP was detected above the selected threshold for chromosome 19. The highest number of significant SNPs was observed for chromosomes 7, 21 and 5, counting 12, 12 and 11 SNPs, respectively. Using the W&C's algorithm negative values were

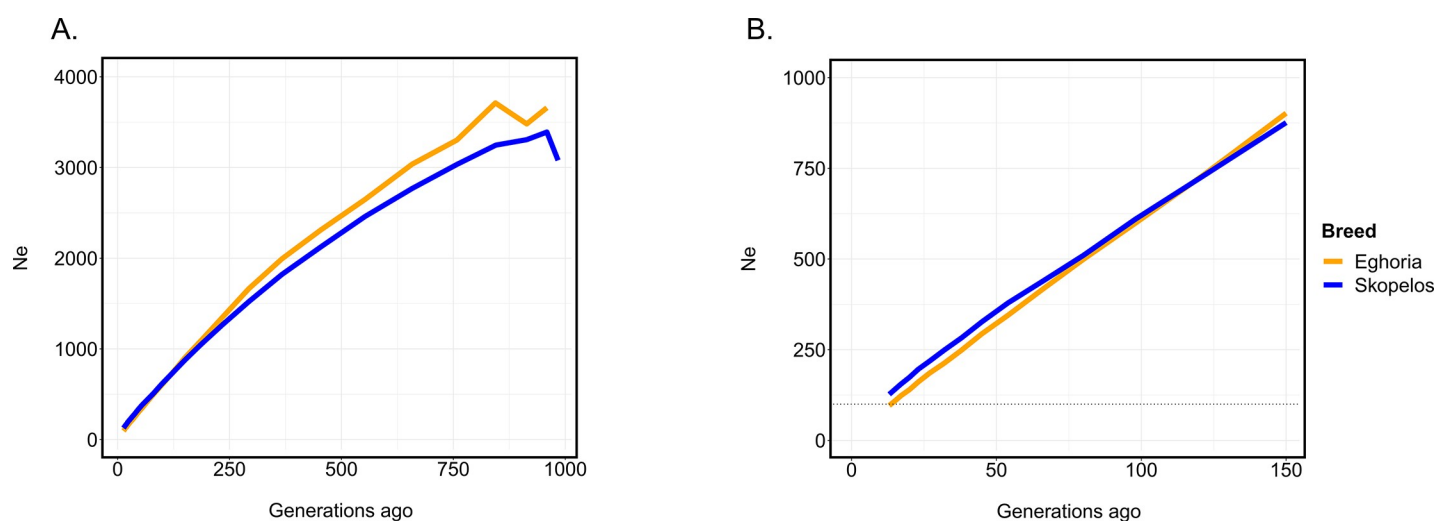


Fig 8. Effective population size (N_e) of Greek goat breeds. Estimation for the two goat breeds was calculated over the last A) 1,000 and B) 150 generations ago. The horizontal dotted line represents $N_e = 100$.

<https://doi.org/10.1371/journal.pone.0226179.g008>

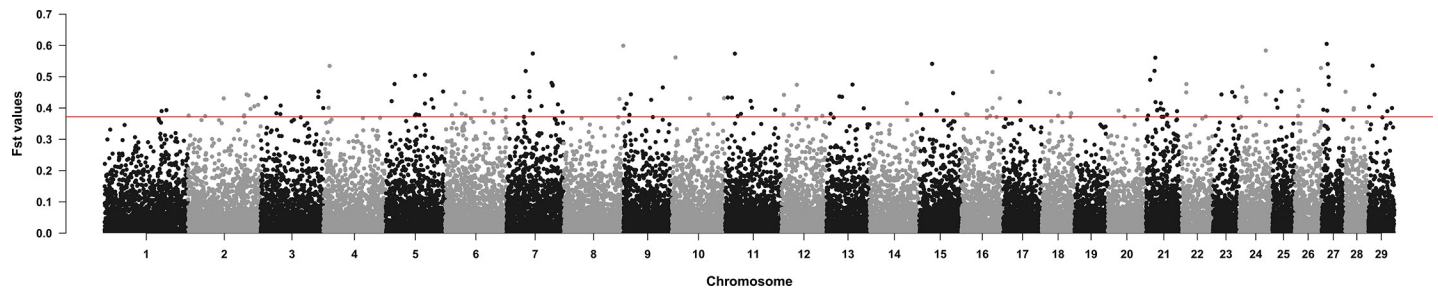


Fig 9. Manhattan plot of the F_{st} values for each single nucleotide polymorphism (SNP). Chromosomes are alternately colored in black and grey. Red line corresponds to the 0.995 of the percentile distribution ($F_{st} = 0.372$).

<https://doi.org/10.1371/journal.pone.0226179.g009>

obtained for 18,724 SNPs, indicating that there is more variation within than between the studied breeds.

The 151 discriminatory SNPs are located within or ± 100 kb of 381 genes. SNPs with the highest F_{st} value were snp30300-scaffold333-3618181, snp17056-scaffold178-6924 and snp7659-scaffold1277-264730, located on chromosomes 27, 8 and 4, respectively. The 151 identified SNPs, alongside with the detected genes in nearby regions are summarized in [S4 Table](#). Interestingly, among the identified genes, a cluster of tRNA genes was found at a higher frequency, and in particular, *TRNAC-GCA* (transfer RNA cysteine (anticodon GCA)) was encountered as the nearest gene of many discriminatory SNPs.

Identification of SNPs for breed assignment using the TRES_all approach and the algorithms implemented in TRES software led to 155 common SNPs among the three methodologies (similarity percentage 77.5%) ([S5 Table](#)). Discriminatory SNPs were located on 28 autosomes; similar to the F_{st} approach, no discriminatory SNP was detected on chromosome 19. Chromosomes 1 and 21 had the largest number of discriminatory SNPs ($N = 13$ SNPs for each chromosome). In total, 389 unique genes were identified within ± 100 kb of the identified SNPs; among them, genes that were located in the nearby regions of the SNPs with the highest F_{st} value, were also present. Algorithms implemented in TRES also indicated that *TRNAC-GCA* was the most frequent neighboring gene of the 155 discriminatory SNPs.

The independent identification of SNPs in the training population (TRES_tt methodology) resulted in 157 SNPs that corresponded to a similarity percentage of 78.5% among the three methodologies implemented in TRES software. SNPs were located in all autosomes; chromosomes 7, 6 and 5 presented the highest number of discriminatory SNPs with 19, 13 and 10 SNPs, respectively. Four-hundred six genes were identified in the nearby regions (within ± 100 kb) of the 157 SNPs ([S6 Table](#)). Evaluation of the 157 SNPs in the test population using GeneClass2 correctly assigned all individuals in the test population.

Common SNPs among the three methodologies were further evaluated in GeneClass2 software. An overlap of discriminatory SNPs per methodology was observed ([S9 Fig](#)). In particular, 95 SNPs (42.6%) were common among the three methods (presented in bold in [S4–S6 Tables](#)). It is noteworthy that W&C's F_{st} and TRES_all methodologies shared 135 discriminatory SNPs (78.9%). Evaluation of the 95 SNPs using different genetic assignment methods correctly assigned all individuals according their origin ([S7 Table](#)). Despite the fact that each individual was correctly assigned to the most likely source population at a strict pre-specified confidence level, assignment probabilities (A.P.) for some animals were low. For the Eghoria breed, these animals originated from farm F4, whereas Skopelos animals with low A.P. originated from farms F3, F5 and F6. Generally, the more admixed an individual was found based on previous analyses (e.g. admixture), the lower A.P. values it presented. Yet, some cases were

observed in which evaluation of individuals using the GeneClass2 software did not agree with admixture analysis. For example, samples SK6, SK10 and SK32 demonstrate identical profile at $K = 3$ in admixture analysis, whereas the A.P. value for Skopelos breed for SK10 is lower compared to SK6 and SK32 animals.

Eighty-four out of the 95 common SNPs were found to be in ROH segments either for Eghoria or Skopelos breeds. The 95 common SNPs were located in or within $\pm 100\text{kb}$ from 237 genes. Among them, some well-characterized genes were included such as *BMP2* (Bone morphogenetic protein 2), *PDGFRB* (b-Type platelet-derived growth factor receptor) and *ZARI* (Zygote arrest 1). The 95 common SNPs were not evenly distributed across autosomes; chromosome 7 had the largest number of identified discriminatory SNPs ($N = 12$), followed by chromosome 21 ($N = 8$) and 5 ($N = 8$).

Discussion

Genetic diversity

In this study the population structure, genetic diversity and inbreeding levels were examined in the two recognized Greek goat breeds, Eghoria and Skopelos, using Illumina's Goat SNP50 BeadChip. Overall, the populations studied here possess high levels of genetic diversity according to the indices used. The use of Goat SNP50 BeadChip identified large numbers of polymorphic loci in both breeds, despite the fact that its design and initial validation was based on other breeds for different breeding purposes (Alpine, Angora, Boer, Creole, Jinlan, Kacang, Saanen, Savanna, Yunling). Calculated heterozygosity values of Greek breeds (mean $H_o = 0.394$, mean $H_e = 0.399$) were in agreement with most published studies employing Goat SNP50 BeadChip. Compared to the average heterozygosity values of 43 European breeds reported by Colli et al., (mean $H_o = 0.369$, mean $H_e = 0.378$) [14], Greek breeds presented slightly higher heterozygosity levels. This was also observed focusing only on breeds reared in the Mediterranean basin such as Spanish [55] or Italian goat breeds [56]. Our heterozygosity results were similar with four goat breeds from Sudan (mean $H_e = 0.400$) [17], the Spanish Bermeya breed (mean $H_e = 0.399$) alongside with Boer (mean $H_e = 0.399$) and Toggenburg (mean $H_e = 0.399$) populations reared in Eastern Africa (Uganda, Kenya, Tanzania) [14]. However, results with such small differences are not significant, especially considering that in a cosmopolitan breed such as Toggenburg, different H_e values have been reported in purebred nuclei, ranging from 0.336 [21] to 0.431 [57] using the same genotyping beadchip. These variations could be strongly dependent on sampling locations. Inconsistencies in heterozygosity results were also observed in Angora purebreds reared in three different continents in which the influence of genetic and geographical isolation as well as different selection patterns showed some discrepancies between the studied populations, mostly indicated by their pairwise F_{ST} values [58]. In our study, the low value of the F_{ST} index ($F_{ST} = 0.04362$) indicates that Greek goat breeds are genetically related based on the SNP dataset used; since values of this magnitude (i.e. close to 0) indicate low differentiation between breeds or that historical gene flow exists between breeds. This result is not in agreement with a previous study on the two autochthonous Greek breeds which generated an F_{ST} value of 0.10000; however, in their study only a small number of SNPs ($n = 26$) was analyzed [5]. To our knowledge, Greek goat breeds have never been genotyped with a genome-wide DNA array; as such, we were not able to directly compare and confirm our results with other datasets.

Population structure

Analysis of Greek goat breeds using 45 breeds reared worldwide revealed that Eghoria and Skopelos cluster together with European breeds, away from African and Asian populations.

The clustering of this continental subdivision, and in particular the grouping of Greek breeds with Italian (JON, RME, GAR, ASP, ARG, MLT in Figs 1 and 2) Carpatian and Spanish breeds (CRP and MUG in Figs 1 and 2, respectively) provides insights into the migration waves from the domestication center to the formation of the present European gene pool. Our findings support the hypothesis reported by Colli et al. [14], which suggested that goats from Central and East Mediterranean regions were influenced by West Africa and Southwest Asia at the same time, thus confirming the role of the Mediterranean basin as a crossroad of post-domestication [14]. However, the obtained bootstrap values from gene-flow analysis for Eghoria and Skopelos breeds in our dataset suggest that another resampling could lead to different reconstruction of their nodes, thus, different population splits and gene flow events.

It is important to highlight that although the advent of genome-wide technologies have created cosmopolitan breeds that are far more productive than the local ones, admixture among Greek and cosmopolitan breeds remains at low levels, despite the fact that in Greece many breeders tend to neglect indigenous, less productive breeds.

Concerning the Greek population structure, our study reveals a geographical pattern, grouping individuals according to the sampling locations. All animals from the northwestern Greece formed a tight cluster without overlaps with clusters of other geographical origin. The tight and distinct cluster observed in individuals from farm 1 reflects the limited gene flow among goat populations in that region due to the presence of geographical barriers (mountains) (S10 Fig). This was also observed in Skopelos breed which is expected since insularization usually involves increased levels of homozygosity as a result of management practices, history and demography [13]. The genetic distance observed between the two Eghoria populations (farms 1 and 4) reflects the large genetic variation that this breed expresses which is mainly related to the colonization of Eghoria purebreds in different geographic areas. The small sample size analyzed in the present study, alongside with the limited number of farm locations used for sampling could influence the results reported here concerning the genetic diversity of this breed. Hence, the observed geographical pattern raises questions regarding the population structure of all the recognized goat types in Greece, especially for the indigenous Eghoria breed. Thus, it would be interesting to explore population structure among populations located in geographically isolated regions and in inaccessible areas or where transhumance is still active.

Population structure analyses revealed that the most admixed population inhabits central Macedonia (northern Greece), which is a region featuring numerous goat farms with different production systems. ADMIXTURE and PCA analyses in farm 4 indicate that admixture between purebreds existed in the past, with exchanges of genetic material between breeds, despite the fact that their phenotypes pointed otherwise. The high levels of admixture in farm 4 in our study are possibly due to an unsupervised and random mating system between animals, which is usual in goat farming in Greece. The grouping of individuals according to the sampling locations evidenced in our results has previously been reported for several goat breeds, where gene flow was studied in their transit from Middle East to the Iberian Peninsula [59, 60].

Patterns of homozygosity

Runs of homozygosity in autosomal SNPs were determined for all samples in each breed, to assess whether regions in the genome are present, that are potentially under positive selection, therefore indicators of selective sweeps in the studied populations [61, 62], or to provide insights on how demography and the recombination landscape have evolved over time for the two breeds [63, 64]. The high levels of genetic diversity in the two breeds, indicated by

heterozygosity levels (H_o , H_e) and Wright's F statistics, were also confirmed by analyzing the degree of inbreeding through indices such as F_{IS} and F_{ROH} . The degree of inbreeding for the two breeds was quite similar in all inbreeding coefficients measured. Since F_{IS} values did not deviate significantly from zero in both breeds, we cannot not claim that Greek breeds are a result of genetic subdivision (Wahlund effect). The same pattern in F_{IS} values is observed in many cosmopolitan goat breeds reared worldwide such as Angora, Alpine, Saanen and Toggenburg [57, 58], as well as in many autochthonous breeds in Italy [56], Sudan [17] and South Africa [65].

Since there is no consensus or standardized criteria to define a ROH within a species, we identified ROH segments in the goat genome, by applying the same parameters used by the AdaptMap project [11, 13]; therefore, our results could be comparable with other breeds reared worldwide. In the present study, inbreeding was also calculated through ROHs, since many studies have shown that F_{ROH} seems to be a more powerful approach for detecting inbreeding than any other alternative method, due to the advantage of not being influenced by estimates of allele frequency or incomplete pedigree data [66–68]. Estimates of pedigree-based inbreeding could not be derived in the present study due to lack of pedigree data in the studied animals. The average fraction of the genome that contains ROH for Eghoria was higher compared to Skopelos breed. Skopelos displayed low F_{ROH} value, similar to breeds from East Africa (Galla, Karamonja, Sonjo, Sebei, Woyito Guji, Gumez), Turkey (Kil, Kilis) and some European breeds from Italy (Argentata), Spain (Malaguena, Bermeya) and Romania (Carpatian) [11, 13]. The increased F_{ROH} value for Eghoria compared to Skopelos breed was equivalent to autochthonous goat breeds from Europe (Valpassiria, Garganica, Ciociara Grigia, Provencale, Alpine) and Africa (Saidi, Burundi, Matebele, Red Sokoto, Guera, Nicastrese, Pare White, Keffa) [11, 13].

According to our results, Greek breeds presented quite smaller number of ROHs per individual but a higher average ROH length compared to various breeds reared worldwide [21] or local breeds [55]. The opposite was true for Swiss goat breeds which presented a quite higher average ROH length (max ROH length = 103.745 Mb) [69]. Interestingly, Eghoria exhibited greater genetic diversity based on the other inbreeding indices but had higher F_{ROH} compared to Skopelos breed. This has been also reported for the Rangeland breed reported by Brito et al. [21] and can be justified by recent selection or inbreeding. Another possible explanation of the increased homozygosity in Eghoria breed could be justified by the consecutive expansions of Eghoria populations over the Greek mainland, thus several founder effects and geographical isolation could be responsible for the increased homozygosity levels. Additionally, many useful SNPs might be omitted from the medium density array used, which could corrupt the ROH segments and lead to different results. True extent of homozygosity can be underestimated due to not clearly defined SNPs, e.g. hemizygous deletions, or due to the fact that sometimes SNPs are not LD-pruned before the data can be used for ROH analysis [70, 71]. Probably, a denser SNP panel, would normalize such inconsistencies and improve prediction accuracies, although it has been documented that increased marker density may improve resolution, but can also decrease power and add noise to the analyses by the use of non-informative SNPs [72].

ROH analysis revealed that many of the detected SNPs within ROHs mapped to genes with well-characterized functions. For both breeds, chromosome 6 demonstrated large homozygous regions, with SNPs within these ROHs mapping to genes affecting milk production traits such as the casein gene cluster containing *CSN1S1* (alpha-S1-casein), *CSN1S2* (alpha-S2-casein), *CSN2* (beta casein) and *CSN3* (kappa casein) [73] or the *ABCG2* (*ATP binding cassette subfamily G member 2*) gene [74]. Moreover, ROH segments were also found to map to the *BMPRII* (bone morphogenetic protein receptor) gene, which is known as a major gene for prolificacy in sheep [75], however its role in goats concerning prolificacy is still unclear and remains to be

explored [76–79]. ROH analysis also revealed regions on chromosome 14, covered by multiple ROH segments for the Eghoria breed (15.2–81.5 Mb); within this region *DGAT* (Diacylglycerol O-Acyltransferase 1) gene is located, which has recently been found to affect milk fat content in dairy goats [80].

Linkage disequilibrium analysis through r^2 confirmed the results obtained from ROH analysis, although differences between breeds were quite modest. In particular, the slightly lower LD decay in Eghoria corroborated increased levels of homozygosity compared to Skopelos breed. Mean r^2 values were lower than other goat breeds [57, 81, 82] but, in general, the pattern of LD decay was a typical curve that is usually observed in livestock species [83–85]. However, setting a threshold for r^2 of 0.3 to ensure successful whole genome association studies [86] and considering the rapid LD decay over short distances [37, 87], we highlight once more the need for a denser SNP panel. Consistent with the results obtained from ROH analysis, the short range of LD proves that Greek breeds differ from the intensively selected ones such as Cashmere, Toggenburg and Boer [81] in which LD decay, as well as their respective effective population sizes, point otherwise.

Despite the differences in their morphological and productive characteristics, Greek breeds seem to share a similar genetic background. Due to breeding practices (intense selection through participation in conservation programs), history and demography of Skopelos breed, we would expect increased levels of homozygosity compared to Eghoria breed. Yet, our results indicated that the two breeds demonstrate similar levels of genetic diversity. Furthermore, accepting 50–100 as a critical effective population size range for long-term viability [88], our results reveal that both Greek goat breeds are of considerable sizes and are not being threatened.

Breed assignment

Discrimination of goat breeds has been studied in the past using microsatellite markers [89]; however, the power of analysis increases when genotyping microarrays are used instead, since associations rely on a much higher number of markers and for this reason microsatellites have now been replaced in the study of breed identification [16, 90, 91]. In our study, by using the Goat SNP50 BeadChip, admixed animals were traced, although their available recorded pedigree data pointed otherwise. Hence, since pedigree data are very limited in Greek indigenous goats, we propose that an alternative way is possible to better assign individuals to their origin, mostly including genomic data. In the present study, we applied three different methods to select a reduced and representative number of markers from Goat SNP50 BeadChip to efficiently discriminate Greek goat breeds.

Population structure and breed assignment using W&C's *Fst* and TRES approaches have been evaluated in goats [12, 21, 69, 92] and avian species [93, 94]. A panel of informative SNPs for parentage assignment in goat breeds reared worldwide have recently been published by the AdaptMap initiative [16], however, no SNP was common with our dataset. In our dataset, algorithms implemented in W&C's *Fst* and TRES methodologies resulted in a high percentage of common discriminatory SNPs. However, further evaluation of the discriminatory SNPs is required, including phenotypic data and subsequent validation in populations from different areas before these SNPs can be efficiently used in animal breeding programs, probably through the design of a customized Greek goat SNP panel. Additional validation of our results with larger population sizes analyzed would be desirable, since the small sample size analyzed in this study could influence the obtained results. With the recent update of the goat genome assembly [9] it is expected that non annotated SNPs will be assigned to genomic locations and associated with biological functions, and the need highlighted by many authors for a denser

genotyping microarray in goats will be fulfilled, making studies on population genetics even more accurate.

An important problem highlighted from the present study is the incorrect classification of purebreds despite their recorded pedigree and morphological characteristics, emphasizing the need for a more effective way to identify and record individuals. Preservation of indigenous purebreds is crucial since the irreversible loss of alleles could lead to decreased adaptability and survival rates especially taking into consideration the threat from the upcoming climatic change [95]. Since goats are among the livestock species that will be directly affected by these changes, well adapted individuals should be exploited to breed the resilient genotypes.

Within $\pm 100\text{kb}$ of the 95 common discriminatory SNPs many candidate genes were found. All the identified genes surrounding the 95 SNPs (237 genes), were extracted and used as a dataset to search against the KEGG database for metabolic pathways relevant to important productive traits (data not shown). Based on this analysis, only general/basic pathways, essential for life maintenance and organism homeostasis were detected e.g. propanoate metabolism, cell cycle, amino sugar and nucleotide sugar metabolism. Among the identified loci, *BMP2* is a well characterized gene, that has been associated in sheep with body size development [92] and tail-type formation [96]. In goats, *BMP2* has been identified as a putative gene involved in the formation and adaptation to the alpine environment of Swiss goat populations [69] and as a regulator of hair follicle development in Cashmere goats [97]. Another gene that was found within the selected window of the discriminatory SNPs was *MYOZ1* (Myozenin 1). This gene has been associated with different lipid traits in cattle [98] and with fetal weight in sheep [99]. In goats, however, its role remains to be explored, although members of the same family (*MYOZ2* and *MYOZ3*) have been studied against their role in meat quality [100]. Moreover, our results indicate that *ZAR1* gene is identified in the nearby regions of discriminatory SNPs, in all methodologies applied. *ZAR1* is a gene well known for its role on embryonic development and fertility control in mammals [101]. Yet, in goats this gene has not been intensively studied, though it was found to be located in a QTL under selection in Saanen goats [21]. *TRNAC-GCA* is a gene reported by many authors that is located around significant or discriminatory SNPs. Specifically, QTLs enclosing this gene have high effect in carcass weight [102] and in mineral (Cr, Fe) concentration in muscles [103] in Nelore cattle. In addition, *TRNAC-GCA* has been related to affect residual feed intake in Angus breed [104] and sperm quality in Holstein-Friesian bulls [105]. In goats, *TRNAC-GCA* has been frequently encountered in genomic regions showing evidence of positive selection for breeds of different purposes [21].

Conclusions

In this study, genetic diversity and population structure of autochthonous Greek goat breeds was assessed for the first time using the Goat SNP50 BeadChip. Our results provide a comprehensive characterization of goat genetic diversity, inbreeding levels and population structure in Greece, using $\sim 60\text{K}$ SNPs. Our findings revealed that the two indigenous goat breeds maintain high levels of genetic diversity which can be further exploited in the design and implementation of breeding schemes. Eghoria presented slightly increased inbreeding levels, nevertheless, our results show that the two breeds are closely related. Moreover, first insights of Greek goat breed traceability are presented, hence, our data could be used for breed discrimination and protection of the authenticity of traditional local products. According to our results, the high levels of genetic diversity in Greek goat breeds could serve as a valuable genetic reservoir to be exploited in national breeding schemes.

Supporting information

S1 Fig. Geographical map indicating the location of goat farms. Sampling for the Eghoria breed was conducted from farm 1 (F1, number of animals (N) = 15) located in Tsepelovo and farm 4 (F4, N = 17) located in Thessaloniki. Sampling for the Skopelos breed was conducted from farm 2 (F2, N = 10), farm 3 (F3, N = 10) and farm 6 (F6, N = 9) located in Skopelos, farm 4 (F4, N = 8) and farm 5 (F5, N = 3) located in Alonnisos island.
(TIFF)

S2 Fig. Fraction of variance explained per migration event (m) using the Treemix software. Each migration event was analyzed in triplicates.
(TIFF)

S3 Fig. Plot of residuals when ten migration edges were fit.
(TIFF)

S4 Fig. Admixture analysis per farm at K = 2 and 3. Each farm is presented by a vertical bar. Different colors indicate different clustering groups. F1: farm 1; F2: farm 2; F3: farm 3; F4: farm 4; F5: farm 5; F6: farm 6.
(TIFF)

S5 Fig. Total number and length (Mbp) of runs of homozygosity (ROHs) per animal. Different colors indicate the two breeds (orange for Eghoria, blue for Skopelos).
(TIFF)

S6 Fig. Distribution of Runs of Homozygosity (ROHs) on chromosomes. Each bar represents the total number of ROHs distributed per chromosome (Number of ROH). The red line shows the average percentage of ROHs for each chromosome.
(TIFF)

S7 Fig. Manhattan plot of the proportion of times (%) SNPs are located within ROHs per breed, plotted against SNP position in autosomes. Autosomes are alternately colored in red and blue. SNPs: Single nucleotide polymorphisms, ROHs: Runs of homozygosity.
(TIFF)

S8 Fig. Linkage disequilibrium (LD) decay over autosomal chromosomes.
(TIFF)

S9 Fig. Venn diagram presenting the common and unique SNPs (Single Nucleotide Polymorphisms) per methodology.
(TIFF)

S10 Fig. Treemix analysis with 1 migration event for the Greek populations.
(TIFF)

S1 Table. Information on farms, sampling locations and number of animals analyzed per breed.
(DOCX)

S2 Table. Quality control of caprine Single Nucleotide Polymorphisms (SNPs).
(DOCX)

S3 Table. Estimates of ancestral effective population size (Ne) over past generations.
(DOCX)

S4 Table. List of the 151 SNPs identified using the Weir and Cockerham's algorithm.

Genes located within $\pm 100\text{kb}$ of the identified SNP or nearby genes of the identified SNPs are presented in *italics*. Texts in bold indicate the 95 common SNPs identified among the three methods. CHR: chromosome, kb: kilo base pair; SNP: Single nucleotide polymorphism. (DOCX)

S5 Table. List of the 155 single nucleotide polymorphisms (SNPs) identified using all animals (TRES_all method). Genes located within $\pm 100\text{kb}$ of the identified SNP or nearby genes of the identified SNPs are presented in *italics*. Texts in bold indicate the 95 common SNPs identified among the three methods. CHR: chromosome, kb: kilo base pair; SNP: Single nucleotide polymorphism. (DOCX)

S6 Table. List of the 157 single nucleotide polymorphisms (SNPs) identified by splitting the dataset into training and test populations (TRES_tt method). Genes located within $\pm 100\text{kb}$ of the identified SNP or nearby genes of the identified SNPs are presented in *italics*. Texts in bold indicate the 95 common SNPs identified among the three methods. CHR: chromosome, kb: kilo base pair; SNP: Single nucleotide polymorphism. (DOCX)

S7 Table. Assignment probabilities for each individual using 95 SNPs, calculated with the GeneClass2 software. A.P.: Assignment Probability, SNPs: Single nucleotide polymorphisms. (DOCX)

Acknowledgments

We thank the Greek General Secretariat for Research and Technology (GSRT) for continuous supporting.

Author Contributions

Conceptualization: G. Th. Tsangaris, A. Tzora, I. Skoufos, G. Arsenos.

Data curation: S. Michailidou, G. Th. Tsangaris, G. Arsenos.

Formal analysis: S. Michailidou.

Funding acquisition: G. Th. Tsangaris, A. Tzora, I. Skoufos, G. Arsenos.

Investigation: S. Michailidou, A. Tzora, I. Skoufos, G. Banos, G. Arsenos.

Methodology: S. Michailidou, G. Arsenos.

Project administration: S. Michailidou, A. Argiriou, G. Arsenos.

Resources: G. Th. Tsangaris, A. Tzora, I. Skoufos, A. Argiriou, G. Arsenos.

Software: S. Michailidou.

Supervision: G. Th. Tsangaris, A. Argiriou, G. Arsenos.

Validation: S. Michailidou, G. Th. Tsangaris, A. Argiriou, G. Arsenos.

Visualization: S. Michailidou, G. Th. Tsangaris, A. Argiriou, G. Arsenos.

Writing – original draft: S. Michailidou, A. Argiriou, G. Arsenos.

Writing – review & editing: S. Michailidou, G. Th. Tsangaris, G. Banos, A. Argiriou, G. Arsenos.

References

1. Food and Agriculture Organization of the United Nations. Accessed 10 Jan 2018 [Internet]. 2014. Available from: <http://www.fao.org/faostat/en/>
2. Gelasakis AI, Rose G, Giannakou R, Valergakis GE, Theodoridis A, Fortomaris P, et al. Typology and characteristics of dairy goat production systems in Greece. *Livest Sci*. 2017; 197:22–9. WOS:000395843100005.
3. Hatziminaoglou Y, Boyazoglu J. The goat in ancient civilisations: from the Fertile Crescent to the Aegean Sea. *Small Ruminant Res*. 2004; 51(2):123–9. WOS:000188873900002.
4. Cappuccio I, Pariset L, Ajmone-Marsan P, Dunner S, Cortes O, Erhardt G, et al. Allele frequencies and diversity parameters of 27 single nucleotide polymorphisms within and across goat breeds. *Molecular Ecology Notes*. 2006; 6(4):992–7. <https://doi.org/10.1111/j.1471-8286.2006.01425.x>
5. Pariset L, Cuteri A, Ligda C, Ajmone-Marsan P, Valentini A, Consortium E. Geographical patterning of sixteen goat breeds from Italy, Albania and Greece assessed by Single Nucleotide Polymorphisms. *BMC ecology*. 2009; 9:20. <https://doi.org/10.1186/1472-6785-9-20> PMID: 19725964; PubMed Central PMCID: PMC2754418.
6. Colli L, Joost S, Negrini R, Nicoloso L, Crepaldi P, Ajmone-Marsan P, et al. Assessing The Spatial Dependence of Adaptive Loci in 43 European and Western Asian Goat Breeds Using AFLP Markers. *Plos One*. 2014; 9(1). ARTN e86668 10.1371/journal.pone.0086668. WOS:000330617100024.
7. Naderi S, Rezaei HR, Taberlet P, Zundel S, Rafat SA, Naghash HR, et al. Large-Scale Mitochondrial DNA Analysis of the Domestic Goat Reveals Six Haplogroups with High Diversity. *Plos One*. 2007; 2(10). ARTN e1012 10.1371/journal.pone.0001012. WOS:000207456000011.
8. Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and characterization of a 52K SNP chip for goats. *PloS one*. 2014; 9(1):e86227. <https://doi.org/10.1371/journal.pone.0086227> PMID: 24465974; PubMed Central PMCID: PMC3899236.
9. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature genetics*. 2017; 49(4):643–50. <https://doi.org/10.1038/ng.3802> PMID: 28263316; PubMed Central PMCID: PMC5909822.
10. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature biotechnology*. 2013; 31(2):135–41. <https://doi.org/10.1038/nbt.2478> PMID: 23263233.
11. Bertolini F, Cardoso TF, Marras G, Nicolazzi EL, Rothschild MF, Amills M, et al. Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genet Sel Evol*. 2018; 50:59. WOS:000450519100005. <https://doi.org/10.1186/s12711-018-0424-8> PMID: 30449279
12. Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, et al. Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet Sel Evol*. 2018; 50(1):57. WOS:000450519100003. <https://doi.org/10.1186/s12711-018-0421-y> PMID: 30449276
13. Cardoso TF, Amills M, Bertolini F, Rothschild M, Marras G, Boink G, et al. Patterns of homozygosity in insular and continental goat breeds. *Genet Sel Evol*. 2018; 50:56. WOS:000450519100002. <https://doi.org/10.1186/s12711-018-0425-7> PMID: 30449277
14. Colli L, Milanesi M, Talenti A, Bertolini F, Chen MH, Crisa A, et al. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet Sel Evol*. 2018; 50(1):58. WOS:000450519100004. <https://doi.org/10.1186/s12711-018-0422-x> PMID: 30449284
15. Stella A, Nicolazzi EL, Van Tassell CP, Rothschild MF, Colli L, Rosen BD, et al. AdaptMap: exploring goat diversity and adaptation. *Genet Sel Evol*. 2018; 50:61. WOS:000451145000001. <https://doi.org/10.1186/s12711-018-0427-5> PMID: 30453882
16. Talenti A, Palhiere I, Tortereau F, Pagnacco G, Stella A, Nicolazzi EL, et al. Functional SNP panel for parentage assessment and assignment in worldwide goat breeds. *Genet Sel Evol*. 2018; 50:55. WOS:000450519100001. <https://doi.org/10.1186/s12711-018-0423-9> PMID: 30449282
17. Rahmatalla SA, Arends D, Reissmann M, Ahmed AS, Wimmers K, Reyer H, et al. Whole genome population genetics analysis of Sudanese goats identifies regions harboring genes associated with major traits. *Bmc Genet*. 2017; 18:1–10. WOS:000413487800001. <https://doi.org/10.1186/s12863-016-0468-0>
18. Boettcher PJ, Tixier-Boichard M, Toro MA, Simianer H, Eding H, Gandini G, et al. Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Anim Genet*. 2010; 41:64–77. WOS:000276775100005.

19. Bruford MW, Ginja C, Hoffmann I, Joost S, Orozco-terWengel P, Alberto FJ, et al. Prospects and challenges for the conservation of farm animal genomic resources, 2015–2025. *Frontiers in genetics*. 2015; 6:314. <https://doi.org/10.3389/fgene.2015.00314> PMID: 26539210; PubMed Central PMCID: PMC4612686.
20. Dalvit C, De Marchi M, Cassandro M. Genetic traceability of livestock products: A review. *Meat Sci*. 2007; 77(4):437–49. <https://doi.org/10.1016/j.meatsci.2007.05.027> PMID: 22061927.
21. Brito LF, Kijas JW, Ventura RV, Sargolzaei M, Porto-Neto LR, Canovas A, et al. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *Bmc Genomics*. 2017; 18:229. WOS:000396758600001. <https://doi.org/10.1186/s12864-017-3610-0> PMID: 28288562
22. Felius M, Theunissen B, Lenstra JA. Conservation of cattle genetic resources: the role of breeds. *J Agr Sci-Cambridge*. 2015; 153(1):152–62. WOS:000347711200014.
23. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2010.
24. Becker RA, Wilks. AR. R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. maps: Draw Geographical Maps. R package version 3.3.0. <https://CRAN.R-project.org/package=maps>.2018.
25. South A. rworldmap: A New R package for Mapping Global Data. *The R Journal* Vol. 3/1: 35–43. 2011.
26. South A. rworldxtra: Country boundaries at high resolution. R package version 1.01. <https://CRAN.R-project.org/package=rworldxtra>. 2012.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. WOS:000249128200012. <https://doi.org/10.1086/519795> PMID: 17701901
28. Schaffner SF. The X chromosome in population genetics. *Nat Rev Genet*. 2004; 5(1):43–51. WOS:000187641100015. <https://doi.org/10.1038/nrg1247> PMID: 14708015
29. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics*. 2011; 12:756. <https://doi.org/10.1038/nrg3098> PMID: 21969038
30. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC bioinformatics*. 2011; 12:246. <https://doi.org/10.1186/1471-2105-12-246> PMID: 21682921; PubMed Central PMCID: PMC3146885.
31. Buchmann R, Hazelhurst S. The ‘Genesis’ Manual. University of the Witwatersrand, Johannesburg. 2015.
32. Milanese M, Capomaccio S, Vajana E, Bomba L, Garcia JF, Ajmone-Marsan P, et al. BITE: an R package for biodiversity analyses. *bioRxiv*. 2017:181610. <https://doi.org/10.1101/181610>
33. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *Plos Genet*. 2012; 8(11). ARTN e100296710.1371/journal.pgen.1002967. WOS:000311891600002.
34. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010; 10(3):564–7. WOS:000276407300020. <https://doi.org/10.1111/j.1755-0998.2010.02847.x> PMID: 21565059
35. VanRaden PM. Efficient Methods to Compute Genomic Predictions. *J Dairy Sci*. 2008; 91(11):4414–23. WOS:000260277200035. <https://doi.org/10.3168/jds.2007-0980> PMID: 18946147
36. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet*. 2011; 88(1):76–82. WOS:000286501500007. <https://doi.org/10.1016/j.ajhg.2010.11.011> PMID: 21167468
37. Al-Mamun HA, Clark SA, Kwan P, Gondro C. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genetics, selection, evolution: GSE*. 2015; 47:90. <https://doi.org/10.1186/s12711-015-0169-6> PMID: 26602211; PubMed Central PMCID: PMC4659207.
38. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet*. 2008; 83(3):359–72. WOS:000259307200007. <https://doi.org/10.1016/j.ajhg.2008.08.007> PMID: 18760389
39. Biscarini F, Cozzi P, Gaspa G, G. M. detectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes. R package version 0.9.5. 2018. <https://CRAN.R-project.org/package=detectRUNS>.
40. Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Al Cavanagh J, Barris W, et al. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel.

- Bmc Genomics. 2008; 9:187. WOS:000256398400001. <https://doi.org/10.1186/1471-2164-9-187> PMID: 18435834
41. Barbato M, Orozco-TerWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet*. 2015; 6:109. WOS:000352769500001. <https://doi.org/10.3389/fgene.2015.00109> PMID: 25852748
42. Corbin LJ, Liu AYH, Bishop SC, Woolliams JA. Estimation of historical effective population size using linkage disequilibria with marker data. *J Anim Breed Genet*. 2012; 129(4):257–70. WOS:000306277300002. <https://doi.org/10.1111/j.1439-0388.2012.01003.x> PMID: 22775258
43. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. 1984; 38(6):1358–70. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x> PMID: 28563791.
44. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*. 2014:005165. <https://doi.org/10.1101/005165>
45. Kavakiotis I, Triantafyllidis A, Ntelidou D, Alexandri P, Megens HJ, Crooijmans RP, et al. TRES: Identification of Discriminatory and Informative SNPs from Population Genomic Data. *J Hered*. 2015; 106(5):672–6. <https://doi.org/10.1093/jhered/esv044> PMID: 26137847.
46. Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, et al. Ethnic-affiliation estimation by use of population-specific DNA markers. *Am J Hum Genet*. 1997; 60(4):957–64. WOS:A1997WT61400026. PMID: 9106543
47. Wright S. The Genetical Structure of Populations. *Ann Eugen*. 1951; 15(4):323–54. WOS: A1951XY09200006.
48. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet*. 2003; 73(6):1402–22. WOS:000187491100015. <https://doi.org/10.1086/380416> PMID: 14631557
49. Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered*. 2004; 95(6):536–9. <https://doi.org/10.1093/jhered/esh074> PMID: 15475402.
50. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. <https://bioinfogp.cnb.csic.es/tools/venny/index.html> 2007–2015.
51. Paetkau D, Slade R, Burden M, Estoup A. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol Ecol*. 2004; 13(1):55–65. WOS:000187005700006. <https://doi.org/10.1046/j.1365-294x.2004.02008.x> PMID: 14653788
52. Paetkau D, Calvert W, Stirling I, Strobeck C. Microsatellite Analysis of Population-Structure in Canadian Polar Bears. *Mol Ecol*. 1995; 4(3):347–54. WOS:A1995RF43700007. <https://doi.org/10.1111/j.1365-294x.1995.tb00227.x> PMID: 7663752
53. Rannala B, Mountain JL. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94(17):9197–201. WOS: A1997XR76500053. <https://doi.org/10.1073/pnas.94.17.9197> PMID: 9256459
54. Mastrangelo S, Portolano B, Di Gerlando R, Ciampolini R, Tolone M, Sardina MT, et al. Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. *Animal: an international journal of animal bioscience*. 2017; 11(7):1107–16. <https://doi.org/10.1017/S1751731116002780> PMID: 28077191.
55. Manunza A, Noce A, Serradilla JM, Goyache F, Martinez A, Capote J, et al. A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. *Genet Sel Evol*. 2016; 48:52. WOS:000381071000001. <https://doi.org/10.1186/s12711-016-0229-6> PMID: 27455838
56. Nicoloso L, Bomba L, Colli L, Negrini R, Milanese M, Mazza R, et al. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet Sel Evol*. 2015; 47:62. WOS:000358985500001. <https://doi.org/10.1186/s12711-015-0140-6> PMID: 26239391
57. Lashmar SF, Visser C, van Marle-Koster E. SNP-based genetic diversity of South African commercial dairy and fiber goat breeds. *Small Ruminant Res*. 2016; 136:65–71. WOS:000374606400010.
58. Visser C, Lashmar SF, Van Marle-Koster E, Poli MA, Allain D. Genetic Diversity and Population Structure in South African, French and Argentinian Angora Goats from Genome-Wide SNP Data. *PloS one*. 2016; 11(5). WOS:000376588600059.
59. Canon J, Garcia D, Garcia-Atance MA, Obexer-Ruff G, Lenstra JA, Ajmone-Marsan P, et al. Geographical partitioning of goat diversity in Europe and the Middle East. *Anim Genet*. 2006; 37(4):327–34. WOS:000239112100004. <https://doi.org/10.1111/j.1365-2052.2006.01461.x> PMID: 16879341
60. Lenstra JA, Tigchelaar J, Biebach I, Hallsson JH, Kantanen J, Nielsen VH, et al. Microsatellite diversity of the Nordic type of goats in relation to breed conservation: how relevant is pure ancestry? *J Anim*

- Breed Genet. 2017; 134(1):78–84. WOS:000393777200010. <https://doi.org/10.1111/jbg.12226> PMID: 27339108
61. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet.* 2012; 91(2):275–92. <https://doi.org/10.1016/j.ajhg.2012.06.014> PMID: 22883143; PubMed Central PMCID: PMC3415543.
 62. Szmatoła T, Gurgul A, Ropka-Molik K, Jasielczuk I, Zabek T, Bugno-Poniewierska M. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. *Livest Sci.* 2016; 188:72–80. WOS:000377729500011.
 63. Peripolli E, Munari DP, Silva M, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet.* 2017; 48(3):255–71. <https://doi.org/10.1111/age.12526> PMID: 27910110.
 64. Bosse M, Megens HJ, Madsen O, Paudel Y, Frantz LA, Schook LB, et al. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* 2012; 8(11):e1003100. <https://doi.org/10.1371/journal.pgen.1003100> PMID: 23209444; PubMed Central PMCID: PMC3510040.
 65. Mdladla K, Dzomba EF, Huson HJ, Muchadeyi FC. Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. *Anim Genet.* 2016; 47(4):471–82. WOS:000379936600007. <https://doi.org/10.1111/age.12442> PMID: 27306145
 66. Keller MC, Visscher PM, Goddard ME. Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics.* 2011; 189(1):237–U920. WOS:000294721600019. <https://doi.org/10.1534/genetics.111.130922> PMID: 21705750
 67. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *Bmc Genet.* 2012; 13:70. WOS:000312139400001. <https://doi.org/10.1186/1471-2156-13-70> PMID: 22888858
 68. Zhang QQ, Calus MPL, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *Bmc Genet.* 2015; 16. WOS:000358122200001.
 69. Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, et al. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Anim Genet.* 2016; 47(6):727–39. WOS:000387359100010. <https://doi.org/10.1111/age.12476> PMID: 27436146
 70. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *Bmc Genomics.* 2011; 12:460. WOS:000295741300001. <https://doi.org/10.1186/1471-2164-12-460> PMID: 21943305
 71. Ku CS, Naidoo N, Teo SM, Pawitan Y. Regions of homozygosity and their impact on complex diseases and traits. *Human genetics.* 2011; 129(1):1–15. <https://doi.org/10.1007/s00439-010-0920-6> WOS:000286396200001. PMID: 21104274
 72. Alvarenga AB, Rovadoski GA, Petrini J, Coutinho LL, Morota G, Spangler ML, et al. Linkage disequilibrium in Brazilian Santa Ines breed, *Ovis aries*. *Sci Rep.* 2018; 8(1):8851. <https://doi.org/10.1038/s41598-018-27259-7> PMID: 29892085; PubMed Central PMCID: PMC5995818.
 73. Selvaggi M, Laudadio V, Dario C, Tufarelli V. Major proteins in goat milk: an updated overview on genetic variability. *Mol Biol Rep.* 2014; 41(2):1035–48. <https://doi.org/10.1007/s11033-013-2949-9> WOS:000330860900052. PMID: 24381104
 74. Wu HJ, Luo J, Wu N, Matand K, Zhang LJ, Han XF, et al. Cloning, sequence and functional analysis of goat ATP-binding cassette transporter G2 (ABCG2). *Mol Biotechnol.* 2008; 39(1):21–7. <https://doi.org/10.1007/s12033-007-9024-5> WOS:000254964700003. PMID: 18256940
 75. Souza CJ, MacDougall C, MacDougall C, Campbell BK, McNeilly AS, Baird DT. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. *The Journal of endocrinology.* 2001; 169(2):R1–6. <https://doi.org/10.1677/joe.0.169r001> PMID: 11312159.
 76. Ahlawat S, Sharma R, Roy M, Mandakmale S, Prakash V, Tania MS. Genotyping of Novel SNPs in BMPR1B, BMP15, and GDF9 Genes for Association with Prolificacy in Seven Indian Goat Breeds. *Anim Biotechnol.* 2016; 27(3):199–207. <https://doi.org/10.1080/10495398.2016.1167706> WOS:000377448200008. PMID: 27135147
 77. Pan ZY, Di R, Tang QQ, Jin HH, Chu MX, Huang DW, et al. Tissue-specific mRNA expression profiles of GDF9, BMP15, and BMPR1B genes in prolific and non-prolific goat breeds. *Czech J Anim Sci.* 2015; 60(10):452–8. <https://doi.org/10.17221/8525-Cjas> WOS:000365982300004.
 78. Polley S, De S, Batabyal S, Kaushik R, Yadav P, Arora JS, et al. Polymorphism of fecundity genes (BMPR1B, BMP15 and GDF9) in the Indian prolific Black Bengal goat. *Small Ruminant Res.* 2009; 85(2–3):122–9. <https://doi.org/10.1016/j.smallrumres.2009.08.004> WOS:000271917600008.

79. Dutta R, Laskar S, Borah P, Kalita D, Das B, Zaman G, et al. Polymorphism and nucleotide sequencing of BMPR1B gene in prolific Assam hill goat. *Mol Biol Rep.* 2014; 41(6):3677–81. <https://doi.org/10.1007/s11033-014-3232-4> WOS:000336404500016. PMID: 24535267
80. Martin P, Palhière I, Maroteau C, Bardou P, Canale-Tabet K, Sarry J, et al. Author Correction: A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. *Sci Rep-Uk.* 2018; 8(1):4060. <https://doi.org/10.1038/s41598-018-22118-x> PMID: 29497092
81. Brito LF, Jafarikia M, Grossi DA, Kijas JW, Porto-Neto LR, Ventura RV, et al. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. *Bmc Genet.* 2015; 16:67. WOS:000356769300001. <https://doi.org/10.1186/s12863-015-0220-1> PMID: 26108536
82. Mucha S, Mrode R, MacLaren-Lee I, Coffey M, Conington J. Estimation of genomic breeding values for milk yield in UK dairy goats. *J Dairy Sci.* 2015; 98(11):8201–8. <https://doi.org/10.3168/jds.2015-9682> PMID: 26342984.
83. Badke YM, Bates RO, Ernst CW, Schwab C, Steibel JP. Estimation of linkage disequilibrium in four US pig breeds. *Bmc Genomics.* 2012; 13:24. <https://doi.org/10.1186/1471-2164-13-24> PMID: 22252454; PubMed Central PMCID: PMC3269977.
84. Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, et al. Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Anim Genet.* 2014; 45(5):754–7. <https://doi.org/10.1111/age.12197> PMID: 25040320.
85. Qanbari S, Pimentel EC, Tetens J, Thaller G, Lichtner P, Sharifi AR, et al. The pattern of linkage disequilibrium in German Holstein cattle. *Anim Genet.* 2010; 41(4):346–56. <https://doi.org/10.1111/j.1365-2052.2009.02011.x> PMID: 20055813.
86. Du FX, Clutter AC, Lohuis MM. Characterizing linkage disequilibrium in pig populations. *International journal of biological sciences.* 2007; 3(3):166–78. <https://doi.org/10.7150/ijbs.3.166> PMID: 17384735; PubMed Central PMCID: PMC1802018.
87. Meadows JRS, Chan EKF, Kijas JW. Linkage disequilibrium compared between five populations of domestic sheep. *Bmc Genet.* 2008; 9:61. Artn 6110.1186/1471-2156-9-61. WOS:000260658300001. <https://doi.org/10.1186/1471-2156-9-61> PMID: 18826649
88. Meuwissen T. Genetic management of small populations: A review. *Acta Agr Scand a-An.* 2009; 59(2):71–9. Pii 91353833410.1080/09064700903118148. WOS:000268709600001.
89. Sardina MT, Tortorici L, Mastrangelo S, Di Gerlando R, Tolone M, Portolano B. Application of micro-satellite markers as potential tools for traceability of Girgentana goat breed dairy products. *Food Res Int.* 2015; 74:115–22. <https://doi.org/10.1016/j.foodres.2015.04.038> WOS:000358971200012. PMID: 28411975
90. Dimauro C, Cellesi M, Steri R, Gaspa G, Sorbolini S, Stella A, et al. Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes. *Anim Genet.* 2013; 44(4):377–82. WOS:000329200100003. <https://doi.org/10.1111/age.12021> PMID: 23347105
91. Ramos AM, Megens HJ, Crooijmans RPMA, Schook LB, Groenen MAM. Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing. *Anim Genet.* 2011; 42(6):613–20. WOS:000296325900006. <https://doi.org/10.1111/j.1365-2052.2011.02198.x> PMID: 22035002
92. Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity.* 2016; 116(3):255–64. WOS:000371736700002. <https://doi.org/10.1038/hdy.2015.94> PMID: 26555032
93. Armstrong C, Richardson DS, Hipperson H, Horsburgh GJ, Küpper C, Percival-Alwyn L, et al. Genomic associations with bill length and disease reveal drift and selection across island bird populations. *Evolution Letters.* 2018; 2(1):22–36. <https://doi.org/10.1002/evl3.38> PMID: 30283662
94. Bosse M, Spurgin LG, Laine VN, Cole EF, Firth JA, Gienapp P, et al. Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science.* 2017; 358(6361):365–8. WOS:000413251000042. <https://doi.org/10.1126/science.aal3298> PMID: 29051380
95. Frankham R. Genetics and extinction. *Biol Conserv.* 2005; 126(2):131–40. <https://doi.org/10.1016/j.biocon.2005.05.002> WOS:000231663000001.
96. Yuan Z, Liu E, Liu Z, Kijas JW, Zhu C, Hu S, et al. Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Anim Genet.* 2017; 48(1):55–66. WOS:000391943100006. <https://doi.org/10.1111/age.12477> PMID: 27807880
97. Su R, Zhang WG, Sharma R, Chang ZL, Yin J, Li JQ. Characterization of BMP2 gene expression in embryonic and adult Inner Mongolia Cashmere goat (*Capra hircus*) hair follicles. *Can J Anim Sci.* 2009; 89(4):457–62. WOS:000274002400004.

98. Dunner S, Sevane N, Garcia D, Leveziel H, Williams JL, Mangin B, et al. Genes involved in muscle lipid composition in 15 European *Bos taurus* breeds. *Anim Genet*. 2013; 44(5):493–501. <https://doi.org/10.1111/age.12044> PMID: 23611291.
99. Xu L, Zhao F, Ren H, Li L, Lu J, Liu J, et al. Co-expression analysis of fetal weight-related genes in ovine skeletal muscle during mid and late fetal development stages. *Int J Biol Sci*. 2014; 10(9):1039–50. <https://doi.org/10.7150/ijbs.9737> PMID: 25285036; PubMed Central PMCID: PMC4183924.
100. Wan L, Ma J, Wang N, Wang D, Xu G. Molecular cloning and characterization of different expression of MYOZ2 and MYOZ3 in Tianfu goat. *PloS one*. 2013; 8(12):e82550. <https://doi.org/10.1371/journal.pone.0082550> PMID: 24367523; PubMed Central PMCID: PMC3867352.
101. Wu XM, Viveiros MM, Eppig JJ, Bai YC, Fitzpatrick SL, Matzuk MM. Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. *Nature genetics*. 2003; 33(2):187–91. WOS:000180773700018. <https://doi.org/10.1038/ng1079> PMID: 12539046
102. Fernandes GA, Costa RB, de Camargo GMF, Carvalheiro R, Rosa GJM, Baldi F, et al. Genome scan for postmortem carcass traits in Nellore cattle. *J Anim Sci*. 2016; 94(10):4087–95. WOS:000388955400004. <https://doi.org/10.2527/jas.2016-0632> PMID: 27898882
103. Tizioto PC, Taylor JF, Decker JE, Gromboni CF, Mudadu MA, Schnabel RD, et al. Detection of quantitative trait loci for mineral content of Nelore longissimus dorsi muscle. *Genet Sel Evol*. 2015; 47:15. WOS:000350951400001. <https://doi.org/10.1186/s12711-014-0083-3> PMID: 25880074
104. Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, et al. Genome-wide association study for feed efficiency and growth traits in US beef cattle. *Bmc Genomics*. 2017; 18(1):386. WOS:000401578600003. <https://doi.org/10.1186/s12864-017-3754-y> PMID: 28521758
105. Kaminski S, Hering DM, Olenski K, Lecewicz M, Kordan W. Genome-wide association study for sperm membrane integrity in frozen-thawed semen of Holstein-Friesian bulls. *Anim Reprod Sci*. 2016; 170:135–40. WOS:000378672900017. <https://doi.org/10.1016/j.anireprosci.2016.05.002> PMID: 27236378